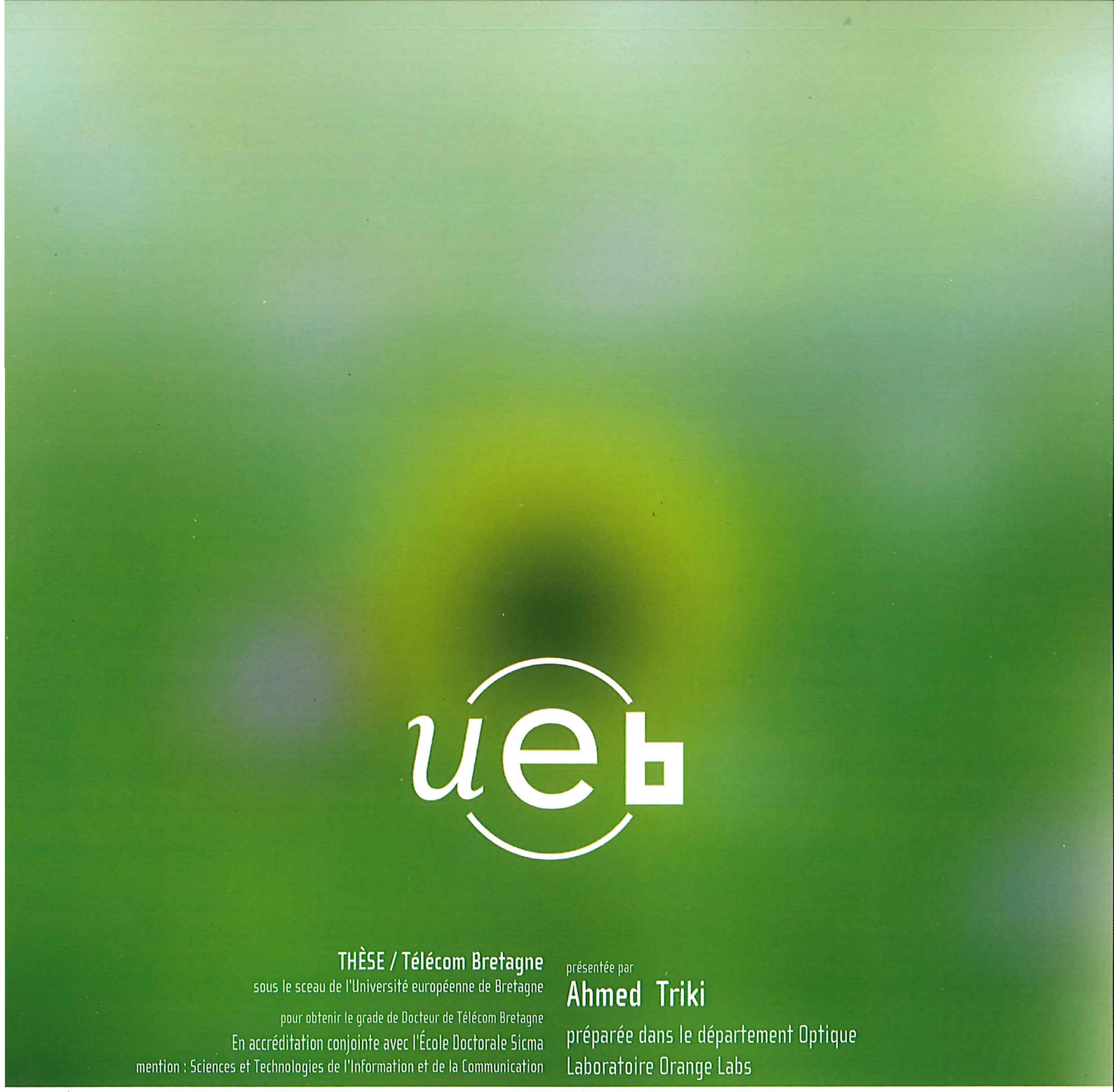
Etude des techniques de transport de données par commutation de rafales optiques sans résolution spectrale de la contention

Ahmed Triki



THÈSE / Télécom Bretagne
sous le sceau de l'Université européenne de Bretagne

pour obtenir le grade de Docteur de Télécom Bretagne
En accréditation conjointe avec l'École Doctorale Sicma
mention : Sciences et Technologies de l'Information et de la Communication

présentée par
Ahmed Triki
préparée dans le département Optique
Laboratoire Orange Labs

Étude des techniques de transport de données par commutation de rafales optiques sans résolution spectrale de la contention

Thèse soutenue le 26 mai 2014
Devant le jury composé de :

Emmanuel Boutillon
Professeur, Université Bretagne-Sud / président

Tülin Atmaca
Professeur, Télécom SudParis / rapporteur

Noureddine Boudriga
Professeur, CNAS University of Carthage – Tunisie / rapporteur

Paulette Gavignet
Ingénieur de recherche, FT/Orange Labs – Lannion / examinatrice

Annie Gravey
Directrice d'études, Télécom Bretagne / examinatrice

Achille Pattavina
Professeur, Politecnico di Milano - Italie / examinateur

Yvan Pointurier
Ingénieur de recherche, Alcatel-Lucent – Nozay / examinateur

Laurent Dupont
Professeur, Télécom Bretagne / directeur de thèse

Bernard Arzur
Ingénieur de recherche, FT/Orange Labs – Lannion / invité

Laurent Bramerie
Ingénieur de recherche, Enssat – Lannion / invité

Philippe Gravey
Directeur d'études, Télécom Bretagne / invité

Esther Le Rouzic
Ingénieur de recherche, FT/Orange Labs - Lannion / invitée

N° d'ordre : 2014telb0317

Sous le sceau de l'Université européenne de Bretagne

Télécom Bretagne

En accréditation conjointe avec l'Ecole Doctorale SICMA

Etude des techniques de transport de données par commutation de rafales optiques sans résolution spectrale de la contention

---

Thèse de Doctorat

Mention: **Sciences et Technologies de l'Information et de la Communication**

Présentée par **Ahmed Triki**

Département : OPTIQUE

Laboratoire : Orange Labs

Directeur de thèse : Laurent Dupont

Thèse soutenue le 26 mai 2014

devant le jury composé de :

M. Emmanuel Boutillon, professeur, Université Bretagne Sud (président)
Mme Tulin Atmaca, professeur, Télécom Sud Paris (rapporteur)
M. Nouredine Boudriga, professeur, CNAS – Tunisie (rapporteur)
Mme Paulette Gavignet, ingénieur de recherche, Orange Labs (examinatrice)
Mme Annie Gravey, professeur, Télécom Bretagne (examinatrice)
M. Achille Pattavina, professeur, Politecnico di Milano (examinateur)
M. Yvan Pointurier, ingénieur de recherche, Alcatel-Lucent (examinateur)
M. Laurent Dupont, professeur, Télécom Bretagne (directeur de thèse)
M. Bernard Arzur, ingénieur de recherche, Orange Labs (invité)
M. Laurent Bramerie, ingénieur de recherche, ENSSAT (invité)
M. Philippe Gravey, directeur d'études, Télécom Bretagne (invité)
Mme Esther Le Rouzic, ingénieur de recherche, Orange Labs (invitée)

• • •

Acknowledgment

Foremost, I would like to thank my Ph.D. examination committee members for the honor that they have given me by accepting to judge this work. Special thanks go to Prof. Atmaca and Prof. Boudriga for their insightful reading of my manuscript and for their valuable and detailed reports.

I sincerely thank my academic supervisors Prof. Laurent Dupont and Philippe Gravey. Their guidance helped me during the different steps of my research work and the writing of my thesis manuscript.

I would like to express my sincere gratitude to my Orange Labs supervisors Paulette Gavignet and Bernard Arzur for the continuous support of my research and Ph.D. study, for their patience, motivation, enthusiasm, and immense knowledge. Throughout these three years, their valuable advice has pushed my research activity in the right direction.

I am thankful to Esther Le Rouzic for the time and effort she has dedicated in reviewing my scientific articles and her constructive comments and remarks. She always found the time for the work that we carried out together.

I would like to express my deepest gratitude to Prof. Annie Gravey. I am grateful for her guidance, direction, support, the time she dedicated to me and invaluable advice along my thesis work. I have enjoyed our discussions and I have learned a lot from her scientific rigor.

I would like to convey my sincere thanks to all the talented researchers who formed part of this work and who helped me to advance my studies. Special thanks to SASER project partners from Telecom Bretagne, INRIA, ENSSAT and Alcatel-Lucent. I would like to take this opportunity to thank my papers' co-authors Edoardo Bonetto, Ramon Aparicio Pardo, Lida Sadaghoon and Olivier Renais for their cooperation and team spirit. I would also like to thank Sebastien Jobert for his significant help and contribution in writing my first patent and for his helpful advice. Special thanks to Jean-Luc Barbey and Thierry Guillossou for their assistance during the design of the TWIN test-bed.

I am grateful to my Orange Labs managers Houmed Ibrahim, Maryse Guena and Françoise Liegeois for their support and encouragement that made this experience enjoyable and rewarding. I am also grateful to my colleagues in TEA/SOAN team for the fruitful exchanges and discussions and for the help they provided during my work among them.

During these past three years, I have met many kind persons and I have made a lot of friends. I would like to sincerely thank my two friends Sofien Blouza and Jelena Pesic. We don't only share the office in Lannion and Brest respectively but we also share great moments together. Their continuous support during these three years has a significant impact on my work. I would like to thank all my friends in Lannion and Brest, especially: Ali, Dhafer, Djamel, Moufida, Narjess, Julie, Jeremy, Mathieu, Nico, Dior, Damien, Mohannad, Sonia, Wafa, Wiliam, Houmid, Hamza, Ion, Bogdan, Zied, Mohamed Tlich and Ahmed Frikha for their friendliness, conviviality and the sympatric moments that we shared together. I have greatly appreciated our talks, our trips and enjoyed our Tennis and Bowling games.

I feel lucky to have friends like you!

I am especially thankful to my grandmother and all my family for the infinite encouragement and the unconditional support. I am indebted to my parents Khaled and Ikram for their love, patience and the effort that they have spent to ensure my education in the best conditions. I would also like to express my gratitude to Amine, Emna, Ameni and Faten to be an inexhaustible source of love and energy.

Résumé

Les réseaux d'opérateur du futur devront supporter des interfaces à haut débit et des besoins dynamiques de bande passante afin de répondre aux augmentations continues de trafic, et à la variabilité de ce dernier. L'introduction de la commutation photonique en sous-longueur d'onde pourrait remplacer avantageusement la commutation électronique de circuit optique ou « Optical Circuit Switching » (OCS) actuellement déployée dans les réseaux opérateur, car cette dernière ne traite efficacement que de gros flux de trafic, puisqu'elle attribue une longueur d'onde par circuit optique. Le concept de commutation en sous-longueur d'onde consiste à partager dynamiquement chaque longueur d'onde entre les flux des différents nœuds de bordure. Ceci requiert la commutation des données optiques en sous-longueur d'onde au niveau des nœuds intermédiaires le long de leurs trajets dans le réseau. Le bénéfice potentiel de la commutation en sous-longueur d'onde dans le domaine optique est qu'en apportant une grande flexibilité à la couche optique, elle permet d'éviter des conversions optique-électrique-optique (O-E-O) coûteuses et gourmandes en énergie, qui sont indispensables lorsque les nœuds intermédiaires commutent les données électroniquement (typiquement, routage IP ou commutation Ethernet).

La commutation en sous-longueur d'onde peut se faire en exploitant la granularité temporelle ou fréquentielle de la longueur d'onde. La commutation de rafales optiques ou « Optical Burst Switching » (OBS) est une technique de commutation en sous-longueur d'onde dans le domaine temporel. Elle a été introduite en 1999 par C.M. Qiao et J.S. Yoo [3] pour compenser le manque de flexibilité de l'OCS et l'immaturité technologique de la commutation de paquets optiques ou « Optical Packet Switching » (OPS). Cette solution a notamment été poussée par la généralisation du trafic IP, se traduisant par un trafic de plus en plus sporadique de nature « paquet ».

L'OBS consiste à regrouper un certain nombre de paquets destinés au même nœud d'extrémité pour former une rafale (ou « burst ») optique. Cette façon de faire permet d'avoir des rafales optiques de durée plus longue que celle des paquets IP ou des trames Ethernet, et ainsi de relâcher les contraintes techniques (vitesse de traitement, durée inter-rafales,...) imposées par la commutation optique des rafales au niveau des nœuds. Comme la durée des rafales optiques reste néanmoins faible (de quelques microsecondes à quelques millisecondes), l'OBS n'est pas trop pénalisé par un délai de transmission excessif. En fait, l'OBS est conçu comme un compromis entre complexité technique et performance.

Outre la capacité des rafales optiques à améliorer la flexibilité des réseaux de transport, on espère des techniques OBS une réduction forte de la consommation électrique de ces réseaux. Cette réduction est justifiée par le fait, qu'idéalement, les rafales optiques sont aiguillées en OBS sans traiter, ni même accéder électroniquement aux données transportées. Par contre, comme les rafales optiques restent dans le domaine optique tout au long de leur trajet du nœud source jusqu'à leurs arrivées au nœud destinataire, des collisions peuvent se produire entre des rafales optiques qui souhaiteraient accéder à un même port de sortie du nœud en même temps. On peut éviter de telles collisions soit de proche en proche, dans chaque nœud, grâce à des stratégies locales qui gèrent les contentions dans les domaines temporel, spectral ou spatial, soit en utilisant une ou plusieurs entités de commande qui réservent les ressources optiques sur la totalité du chemin suivi par les rafales.

Depuis l'apparition du concept de commutation de rafales optiques, plusieurs solutions et mécanismes ont été proposés [68][72][81][84]. TWIN (Time-domain Wavelength Interleaved Networking) est l'une de ces solutions. TWIN évite les pertes de rafales, avec des nœuds

intermédiaires passifs fonctionnant uniquement dans la couche optique, sur une topologie pouvant être maillée. L'avantage majeur de cette solution est que tous les traitements et les processus électroniques se font dans les nœuds périphériques (quand les données à transporter sont encore dans le domaine électrique) avec des nœuds intermédiaires qui sont totalement passifs, ce qui permet d'éliminer les conversions O-E-O et de réduire ainsi la consommation électrique de ces nœuds.

Pour mettre en évidence l'efficacité énergétique de TWIN, nous avons mené une étude énergétique préliminaire consistant à déterminer le nombre de transpondeurs optiques requis pour différentes solutions de transport optique. Les technologies étudiées ont été classées en deux grandes catégories : des technologies basées sur la commutation de circuit et des technologies basées sur la commutation en sous-longueur d'onde. Les technologies à commutation de circuit sont l'opaque, le transparent et l'hybride [62]. Les technologies à commutation de sous-longueur d'onde sont le L-OBS (Label Optical Burst Switching), TWIN et le POADM (Packet Optical Add Drop Multiplexer) [72]. Notre étude, présentée dans chapitre 4 de ce rapport, montre qu'à faible et moyen débit, TWIN et POADM nécessitent moins de transpondeurs que les autres technologies. A haut débit, le nombre de transpondeurs requis par ces deux technologies devient proche de celui demandé par la solution hybride. A ce niveau l'efficacité des plans de commande joue un rôle déterminant sur la connaissance des nombres de transpondeurs requis.

Dans un réseau TWIN, chaque longueur d'onde est dédiée au transport des rafales de données vers un unique nœud destinataire. La transmission des rafales optiques est effectuée selon une structure d'arbre associé à chaque longueur d'onde et dont le nœud destinataire représente la racine et les nœuds source représentent les feuilles. Le nœud source TWIN comprend un (ou plusieurs) transmetteur laser accordable(s) lui permettant d'envoyer les rafales optiques aux différentes destinations, tandis que le nœud destinataire dispose d'un récepteur fixe recevant les rafales sur la longueur d'onde qui lui est attribuée. Les nœuds intermédiaires entre la source et la destination agrègent optiquement les rafales optiques provenant des feuilles et l'aiguillent vers la racine sans aucun traitement électronique : TWIN est basé sur un routage en longueur d'onde, passif et transparent au niveau des nœuds intermédiaire. Comme les feuilles partagent la même longueur d'onde vers la destination, des collisions entre les rafales optiques peuvent survenir et doivent être évitées grâce au plan de commande du réseau. Ainsi, la simplicité photonique des nœuds TWIN impose de recourir à un plan de commande dont le rôle principal est d'éviter les collisions entre rafales optiques destinées à une même feuille, tout en permettant à chaque source d'utiliser son, ou ses transmetteurs laser efficacement.

Le plan de commande doit donc gérer les émissions des rafales optiques, au niveau des sources, de telle sorte que les collisions entre rafales optiques soient évitées au niveau des nœuds intermédiaires et que les blocages de l'émetteur soient réduits au minimum. Le plan de commande est supporté par un réseau séparé du plan de données et qui relie tous les nœuds.

Nous avons proposé dans le que l'émission et la réception des rafales optiques soient organisées selon des cycles successifs appelés « cycles de commande». La durée des cycles de commande est commune à toutes les destinations et elle est supérieure au temps d'aller-retour maximum observé entre l'entité de commande et les sources. Le cycle de commande est composé d'un nombre prédéterminé de « cycles de données ». Le cycle de données est divisé en slots temporels. Le slot temporel permet de transporter une unique rafale sur chaque longueur d'onde et représente donc la granularité la plus fine d'allocation des ressources. Les cycles de données appartenant au même cycle de contrôle présentent la même configuration d'allocation des ressources qui indique comment les ressources optiques (c'est à dire les slots

portés par toutes les longueurs d'onde) sont réparties entre les flux. Cette configuration change d'un cycle de commande à un autre selon les évolutions de trafic.

Dans le chapitre 4, nous avons décrit comment l'allocation de ressources peut évoluer dynamiquement en fonction des évolutions du trafic. Pour assurer la prise en compte de cette évolution, chaque source estime ses besoins en ressources pendant un cycle de contrôle, et communique cette estimation à l'entité de contrôle grâce à un message de « requête » (REQUEST). L'entité de contrôle collecte les besoins de toutes les sources, et calcule localement les nouvelles configurations qui sont communiquées aux sources sous forme d'un message « permission » (GRANT).

L'algorithme local d'allocation des ressources attribue les slots aux flux selon les requêtes des sources en prenant en considération les contraintes suivantes :

1. les rafales optiques ne doivent pas subir de collisions tout au long de leur chemin de la source vers la destination ;
2. le transmetteur ne peut envoyer qu'une seule rafale optique à la fois (éviter le phénomène de blocage) ;
3. le récepteur ne peut recevoir qu'une seule rafale optique à la fois.

La structure en arbre caractérisant la transmission dans un réseau TWIN rend la première contrainte superflue, puisque le fait d'éviter la collision des rafales optiques au niveau de la destination (troisième contrainte) évite implicitement les collisions au niveau de tous les nœuds intermédiaires (première contrainte).

Dans notre thèse, nous proposons plusieurs mécanismes pour les plans de commande, de gestion et de données. Les mécanismes concernant le plan de commande et de gestion dépendent en particulier de la localisation de l'entité qui calcule l'allocation des ressources (centralisée ou distribuée), de la réactivité du plan de commande (statique ou dynamique) et de la topologie du réseau de donnée (avec ou sans point de passage obligé). Les mécanismes concernant le plan de données sont principalement liés à la manière dont les slots sont utilisés (séparés ou fusionnés), la répartition temporelle et la différentiation des classes de services (CoS).

Dans ce cadre, nous avons proposé deux types de plans de commande : centralisé et distribué. Dans l'approche centralisée, une seule entité de commande gère toutes les réservations en attribuant à chaque nœud source les slots à utiliser pour transmettre ses rafales optiques à une destination donnée. L'avantage de cette approche est qu'elle améliore l'utilisation de la bande passante puisqu'elle peut mettre en œuvre une optimisation globale de cette utilisation. En contrepartie, la complexité des algorithmes d'optimisation au niveau de l'entité de commande et la latence provoquée par ce processus centralisé peuvent présentent des inconvénients significatifs. Dans l'approche distribuée, chaque destination va directement contrôler les temps d'émissions des sources qui lui envoient du trafic. Cette approche réduit la complexité du processus de commande puisque chaque destination gère un nombre restreint de réservations. Par contre, la source peut recevoir des autorisations issues de destinations différentes l'enjoignant à émettre du trafic simultanément vers plusieurs destinations, ce qui est impossible ; la source devra donc sélectionner une unique destination pour chacun des slots temporels où un tel conflit existe, ce qui risque de ne pas lui permettre de servir tout le trafic à émettre. Afin de comparer ces deux types de plans de commande, nous avons effectué des simulations en utilisant l'outil OMNET++. La comparaison est effectuée en termes de : délai de bout en bout, gigue, longueur des files d'attente et taux d'utilisation des longueurs d'onde. En ce qui concerne l'allocation des slots, nous avons considéré deux options possibles : une allocation « contigüe » (groupant autant que possible les slots alloués à un flux source-destination donné) et une allocation « disjointe » qui répartit les slots alloués dans le

cycle de données. Les simulations ont montré qu'un plan de commande centralisé est plus performant qu'un plan distribué et en particulier, qu'un plan centralisé avec une allocation contigüe permet d'allouer environ 15% plus de ressources qu'un plan de commande distribué.

Sur la base de ces résultats, nous avons alors comparé la performance de plusieurs algorithmes centralisés. Dans cette étude, nous différencions deux types d'algorithme : « statique » et « dynamique ». L'algorithme « statique » suppose une connaissance préalable de la matrice de trafic ; il est basé sur l'optimisation globale de l'allocation des ressources qui définit pour une longue période de temps l'allocation des slots dans le cycle de données (ici la durée du cycle de commande est de quelques secondes à plusieurs minutes). En opposition à l'algorithme statique, les algorithmes dynamiques changent les allocations en fonction de la variation du trafic observée à courte durée (un cycle de commande durant alors quelques millisecondes seulement). Ils sont basés sur des approches heuristiques réalisant l'allocation des ressources (i.e. le calcul de l'ordonnanceur). Dans ce contexte, nous avons proposé trois algorithmes dynamiques : « disjoint », « contigu » et « hybride ». L'allocation hybride divise la bande passante en deux parties : les ressources de la première partie sont allouées statiquement, tandis que les ressources de la deuxième partie sont allouées dynamiquement. Les performances de chaque algorithme ont été évaluées, par simulation en utilisant l'outil OMNET++, en termes de : délai de bout-en-bout, temps d'attente, temps de service, gigue, débit et taux d'utilisation des ressources. Les résultats obtenus, en considérant un profil synthétique du trafic durant les simulations, montrent que le schéma statique est plus performant que les schémas dynamiques ou hybrides, et permet une utilisation de bande-passante de plus que 80%.

Pour confirmer ces résultats et vérifier la robustesse des schémas statiques, nous avons ensuite utilisé des traces de trafic réel pour alimenter les simulations. A notre connaissance, c'est l'une des rares contributions dans ce type d'étude où des traces réelles sont utilisées au lieu de modèles synthétiques.

Dans cette étude présentée dans le chapitre 5, nous avons proposé d'appliquer TWIN à une architecture « MEET » (Multi-hEad sub-wavElength swiTching) destinée à remplacer l'architecture actuelle des réseaux metro-backhaul. MEET permet de relier la zone métropolitaine avec la zone cœur du réseau opérateur. Elle autorise la communication directe entre les nœuds de ces zones en évitant de passer par un « nœud de concentration » qui contrôle aujourd'hui l'échange de trafic entre la zone métropolitaine et le réseau cœur. Dans l'architecture MEET, « les nœuds d'extrémité » au niveau métro et cœur sont supportés par des nœuds TWIN et « le nœud de concentration » est remplacé par un nœud intermédiaire TWIN tout optique. MEET permet d'une part d'« aplatir » l'architecture actuelle de type « hub-and-spoke » et de remplacer d'autre part des étages d'agrégation électrique par de l'agrégation optique, potentiellement économe en énergie. Nous avons proposé deux alternatives pour le plan de commande à appliquer à MEET: l'approche dynamique basée sur l'heuristique contiguë et l'approche statique basée sur l'optimisation globale. Nous avons proposé plusieurs options d'assemblage de la rafale optique dans le plan de données. En ce qui concerne la taille de la rafale optique, soit nous utilisons des rafales de même taille, en respectant systématiquement un temps de garde entre rafales adjacentes (mode « uni-slotté), soit nous considérons les slots adjacents destinés à un même flux comme un unique intervalle temporel appartenant à une seule rafale (sans temps de garde) ce qui conduit à construire des rafales optiques de taille variable (mode multi-slotté). En ce qui concerne la priorisation entre types de trafic, soit les paquets de données sont insérés dans les rafales en mode FIFO, indépendamment de leur classe de trafic, soit les paquets sont insérés dans les rafales en servant prioritairement les paquets ayant des besoins en matière de gigue et de délai. L'ordre des paquets dans leurs flux respectifs est toutefois maintenu.

Nous avons mené une étude de performance par simulation en utilisant l'outil OMNET++. Les résultats montrent que malgré la forte variation du trafic réel, le plan de commande statique, couplé avec la méthode « multi-slotté », permet de satisfaire les exigences en qualité de service dans les réseaux métropolitains, même à haut débit. Les résultats montrent aussi que la prise en considération des classes de service dans l'assemblage conduit à de meilleures performances pour le trafic sensible au délai au prix de la dégradation de la QoS des flux de moindre priorité.

Enfin, dans le chapitre 6, nous rapportons les étapes de conception d'un banc expérimental qui doit permettre de mieux comprendre les contraintes technologiques de la commutation en sous-longueur d'onde et plus particulièrement de TWIN. Nous avons donc conçu et mis en œuvre un banc de test pour TWIN utilisant un plan de commande centralisé basé sur l'approche statique. La topologie choisie pour ce banc expérimental est composée de quatre nœuds périphériques (deux nœuds sources et deux nœuds destinataires) et un nœud cœur. Les deux nœuds sources sont gérés par une unique entité de commande. L'entité de commande envoie périodiquement des messages « grant » aux sources pour les informer de la configuration d'émission qu'elles doivent appliquer. Ce module est développé en utilisant l'outil LabView Real Time de National Instruments. Le nœud source est constitué d'une unité de contrôle et une unité de transmission de rafale optique. L'unité de commande assure la communication avec l'entité de commande et elle est développée en utilisant l'outil LabView FPGA. Tandis que l'unité de transmission de rafale optique assure l'émission de la rafale optique selon la configuration d'émission proposée par l'entité de contrôle. Cette unité est optoélectronique constituée principalement d'un contrôleur, d'un générateur de rafales électriques, d'une unité de synchronisation externe, d'un laser accordable et d'un modulateur externe. Le nœud destinataire est composé d'une photodiode qui joue le rôle d'une unité de réception de rafales optiques. Grâce à ce banc de test, nous avons montré que le développement d'un nœud TWIN avec un plan de commande statique est actuellement faisable avec les paramètres suivants : rafale optique de 4.5 µs, temps de garde de 0.5 µs et débit de lien égal 10 Gbps. Malgré l'insuffisance de certains composants, nous avons réussi à assurer la synchronisation entre les différentes parties du banc et à obtenir l'exactitude temporelle souhaitée avec des signaux optiques de bonne qualité.

A travers ce travail, nous avons étudié la technologie TWIN en se focalisant sur son plan de commande pour gérer efficacement les ressources optiques. Cette étude a été menée selon de multiples axes : théorique, architecturale et expérimentale.

Comme prochaine étape, nous comptons également étudier le potentiel des solutions de commutation sous-longueur d'onde, notamment TWIN, pour faire face aux nouvelles technologies de transport optique qui ont gagné un grand élan comme les réseaux Flexgrid. Nous comptons aussi explorer la compatibilité des réseaux SDN (Software Defined Network) avec TWIN pour mettre en œuvre un plan de contrôle flexible.

Abstract

Future networks will have to support very high bitrate interfaces and to ensure dynamic bandwidth provisioning in order to deal with increasing and time-varying traffic demands. In this context, a sub-wavelength switching paradigm may be more appropriate than the currently deployed Optical Circuit Switching (OCS) as it brings flexibility in the optical layer, while consuming less energy than electronic switching. Sub-wavelength optical switching consists in dynamically sharing a given wavelength between several source-destination pairs in the optical domain. This requires switching "optical bursts" at the intermediate nodes in the network (i.e. Optical Burst Switching, OBS).

Time-domain Wavelength Interleaved Networking (TWIN) is a promising OBS solution. It consists in allocating a wavelength per destination, and in scheduling all traffic for this destination on a multipoint-to-point tree between sources and this destination. Each source requires a tunable transceiver, whereas a destination only requires a fixed receiver. TWIN has been proposed by Bell Labs in 2003, and has been shown to provide lossless OBS with simple and optically transparent intermediate nodes, within a mesh network. However, TWIN requests a complex control plane in order to avoid burst contention.

Through this thesis, we revisit the original proposal of the TWIN architecture and mechanisms and propose several algorithms to realize the management/control plane and the data plane for a TWIN network on a metropolitan area topology.

We consider either dynamic control planes that realize a closed loop control avoiding burst contention on the basis of a dynamic evaluation of requested resources, or a static control plane that operates in open loop, under the assumption that requested resources are known (e.g. thanks to management plane information). The dynamic control planes are based on a heuristic approach for resource allocation, which changes according to the traffic variation observed during a short period (a "control cycle" of several milliseconds duration). On the other hand, the static scheme is based on an optimized resource allocation implementing an Integer Linear Programming (ILP) formulation.

We first compare dynamic centralized and distributed versions of the control plane in terms of end-to-end delay, jitter, queue length and bandwidth utilization. We then compare the performance of different centralized control planes that can be either dynamic or static. The results obtained by considering synthetic (Poisson) traffic profiles during the simulations show that the static scheme performs better than all dynamic schemes.

In order to confirm these preliminary results and verify the robustness of the static scheme, we have used a real traffic trace to drive the next set of simulations. We have proposed to apply TWIN to a new architecture, MEET, which is intended for a metro-backhaul network. This architecture limits the number of electrical aggregation stages between metro and core networks; it also allows supporting both "hub-and-spoke" and "any-to-any" architectures. Results show that, despite the high variation of the actual traffic, the static scheme still performs well, and better than the dynamic schemes. We also prove that coupling the centralized control plane of MEET with a QoS-aware burst assembly mechanism allows to satisfy multiple classes of service, and to increase network efficiency.

Lastly, we prove the feasibility of TWIN by designing an experimental node operating with a static control plane. On a small test-bed, we have succeeded in ensuring synchronization and in obtaining correct time accuracy with good quality optical signals.

Table of Contents

List of Figures

List of Tables

List of Symbols

AS	Autonomous System
ATM	Asynchronous Transfer Mode
BPG	Burst Pattern Generator
CAPEX	CAPital EXpenditure
CBR	Constant Bit Rate
CE	Control Entity
CU	Control Unit
CET	Carrier Ethernet Transport
CN	Concentration Node
CO	Central Office
C-OBS	Conventional Optical Burst Switching
CoS	Class of Service
CSMA-CD	Carrier Sense Multiple Access with Collision Detection
DBA	Dynamic Bandwidth Allocation
DSL	Digital Subscriber Line
DSLAM	Digital Subscriber Line Access Multiplexer
DWDM	Dense Wavelength Division multiplexing
EDFA	Erbium Doped Fiber Amplifier
EDGE	Enhanced Data Rates for GSM Evolution
EPON	Ethernet Passive Optical Network
EPS	Evolved Packet System
E-UTRAN	Evolved UMTS Terrestrial Radio Access Network
EXC	Electrical Cross Connect
FEC	Forward Error Code
FPGA	Field Programmable Gate Array
FTTB	Fiber To The Building
FTTC	Fiber To The Curb
FTTH	Fiber To The Home
GERAN	GSM EDGE Radio Access Network
GFP	Genetic Framing Protocol
GMPLS	Generalized Multi-Protocol Label Switching
GPON	Gigabit Passive Optical Network
GPRS	General Packet Radio Service
GPS	Global Positioning System
GSM	Global System for Mobile telecommunication

HDLC	High-level Data Link Control
HE	Head End
HFC	Hybrid Fiber Coax
HSPA	High Speed Packet Access
IBP	Internet Backbone Provider
IN	Internet Node
I/O	Input/Output
IP	Internet Protocol
ISP	Internet Service Provider
ITU	International Telecommunication Union
IXP	IP interconnection Point
L-OBS	Labelled Optical Burst Switching
MAC	Media Access Control
MAN	Metro Area Network
MEET	Multi-hEad sub-wavElength swiTching
MII	Media Independent Interface
MN	Multiservice Node
MPLS	Multi-Protocol Label Switching
MTU	Maximum Transmission Unit
NFS	Network File System
NI	National Instruments
NIU	Network Interface Unit
OAM	Operations, Administration and Management
OBS	Optical Burst Switching
OBTN	Optical Burst Transport Network
OCS	Optical Circuit Switching
O-E-O	Optical to Electrical to Optical
OLT	Optical Line Termination
ONU	Optical Network Unit
OPEX	OPerational Expenditure
OPS	Optical Packet Switching
OPST	Optical Packet Switching and Transport
OSA	Optical Spectrum Analyzer
OSI	Open Systems Interconnection model
OTI	Open Transit Internet
OTN	Optical Transport Network
OTT	Over-The-Top

OUT	Optical channel Transport Unit
OXC	Optical Cross Connect
PDH	Plesiochronous Digital Hierarchy
PDN	Packet Data Network
PLMN	Public Land Mobile Network
POADM	Packet Optical Add/Drop Multiplexing
PoC	Proof of Concept
PON	Passive Optical Network
PoP	Point of Presence
PPP	Point-to-Point Protocol
PSTN	Public Switched Telephone Network
QoS	Quality of Service
RBCI	Réseau Backbone de Collecte IP
RF	Radio Frequency
RN	Regional Node
ROADM	Reconfigurable Optical Add Drop Multiplexer
RR	Round Robin
SDH	Synchronous Digital Hierarchy
SDN	Software Defined Network
SFD	Start Frame Delineation
SLPSN	Sub-Lambda Photonically Switched Networks
SNMP	Simple Network Management Protocol
SOA	Semiconductor Optical Amplifier
TCM	Tandem Connection Monitoring
TCP	Transmission Control Protocol
TDM	Time Division Multiplexing
TN	Transit Node
ToS	Type of Service
TPON	Telephony over Passive Optical Networks
TWIN	Time-domain Wavelength Interleaved Networking
UDP	User Datagram Protocol
UMTS	Universal Mobile telecommunication System
UTRAN	Universal Terrestrial Radio Access Network
VoD	Video on Demand
VOQ	Virtual Output Queue
WAN	Wide Area Network
WDM	Wavelength Division Multiplexing

WR-OBS	Wavelength-Routed Optical Burst Switching
WSS	Wavelength Selective Switch

Chapitre I. **Introduction**

The increasing number of connected end user terminals (smartphones, tablets, connected TV) and the race towards higher definition contents contribute to increase the traffic load of telecommunication networks. Consequently, the associated IP traffic growth, mainly driven by audiovisual consumption, challenges the way operators build and operate their network. Content Delivery Networking (CDN) reduces the load of core transport network but, it does not help in optimizing backhaul network architecture which has to support the growing traffic load. Transparent caching helps handling Over The Top (OTT) content which represent the largest part of the traffic, but this solution does not seem long-lasting facing OTT content distribution strategies (URL redirection, content encryption). In the specific situation where the operator is involved neither in the control nor the distribution of the content, operators investigate opportunities to improve the network scalability while reducing both capital and operational expenditures (CAPEX and OPEX). Thus, the traditional transport layers technology using Time Division Multiplexing (PDH/SDH/OTN) [1] is gradually replaced by technologies that better fit with current needs. More flexibility is also required to take into account the modification of the traditional concentration/distribution traffic pattern in the backhaul network into a more general mesh pattern. A possible solution is to improve the flexibility of the optical transport layer so that it becomes compliant with packets networks now predominant in both residential and business services.

The new transport approach should support very high bitrate interfaces, dynamic bandwidth provisioning such that it meets customer's demands and guaranteed quality of service (QoS), and uses the existing fibers. In this context, a sub-wavelength switching paradigm may be more appropriate than the currently deployed Optical Circuit Switching (OCS) to bring the required flexibility in the transport layer.

Optical Packet Switching (OPS) [2] is the most ambitious solution of sub-wavelength switching paradigm. It allows switching packets optically with time duration in the range of

10 ns to 1 µs. The information used to switch packets at the optical level is carried in the header. The main technological issue with such approach is the lack of mature optical memory and fast pure optical switches for which no industrial solution is foreseen before 2020. Optical Burst Switching (OBS) [3] was introduced in 1999 as an alternative to OPS. OBS assembles bursts of packets intended to the same destination such that the duration of the obtained optical bursts is in the range of 1 µs to 10 ms. This relaxes the constraints on optical switching functions and provides a compromise between performance and technical complexity. Using OBS in the backhaul could help in improving network efficiency by offering the ability to overcome the coarse granularity of OCS that provides optical switching by allocating one wavelength channel to each source-destination pair. OBS could also avoid the high cost and delays resulting from Optical-Electrical-Optical (O|E|O) converters being deployed at all the optical switches. The traffic in transit in the OBS node stays in the optical layer and does not need to be switched electronically. This could become a significant advantage in a context where traffic changes unpredictably and flow distributions can evolve according to OTT content distribution strategies.

Many declinations and variants derived from the original OBS concept have been proposed in the literature. The Sub-Lambda Photonically Switched Networks (SLPSN) [4] is the recently proposed term in some ITU contributions for this concept.

Time-domain Wavelength Interleaved Networking (TWIN) [5] is among the interesting SLPSN solutions that can support mesh topology. TWIN has been proposed by a group of Bell Labs researchers; is an optics-based transport network architecture that aims to provide optical grooming without burst loss. It uses a wavelength routing approach based on the pre-configuration of so-called “light trees”. Each light tree uses a specific wavelength channel and is associated to a unique destination node. In TWIN, a source node selects its destination by transmitting on the corresponding wavelength. Burst collision in the tree’s merge points is avoided via scheduling performed by the control plane. Through this thesis, we carry out studies on the SLPSN solutions. Specifically, we study the TWIN approach and we analyze how the control plane can couple scheduling to traffic measurement. Then, we use this concept to propose a new architecture that extends the metro-backhaul network to optically reach remote core nodes, leading to the elimination of some electrical aggregation stages.

Finally, we perform an experimental study implementing TWIN in order to demonstrate the feasibility of this concept. The present report is organized as follows:

Chapter 2 focuses on describing the context of sub-wavelength switching technologies from both architectural and protocol perspectives. The goal of this study is to identify the potential use cases suitable to deploy the OBS technology. Therefore, we first describe realistic Internet Service Provider's (ISP) network architecture. We consider three separate levels: access, metro-backhaul and core levels. We illustrate this first section by examples from the network of the French operator "Orange". We then focus on the protocol stacks deployed in a metro-core network. We emphasize in particular the transport protocols and the medium access techniques.

In chapter 3 we present the sub-wavelength switching solutions currently described in the literature. These solutions are to be applied in a metro-core network. We consider some criteria related to topology, type of the control plane, synchronization issues, etc. Nevertheless, the most relevant criterion that we use to classify these solutions is based on whether bursts are potentially lost along their trip within the network. For "lossy" solutions, congestion can be resolved in the time domain (delaying), the spatial domain (deflection), and/or the spectral domain (wavelength conversion). However, none of these methods is really efficient to yield an acceptable burst loss ratio in conjunction with correct latency. On the other hand, "lossless" solutions appear as viable to reach an acceptable throughput while keeping a reasonable latency. Thus, they fit better to the operator needs than lossy approaches. Lossless solutions generally rely on a sophisticated control plane to prevent burst contention; they rely on schedules being centrally computed and distributed to distant nodes, which necessitates a tight synchronization between nodes. In order to address this particular issue, we have submitted a patent about a solution to perform the synchronization in a TWIN distributed control plane.

In chapter 4, we carry out an in-depth study on the TWIN paradigm. We first compare the number of required transponders for some optical sub-wavelength switching solutions and for legacy circuit switching solutions. Results show that sub-wavelength switching technologies can reduce the number of transponders, which implies that these technologies could be promising solutions to reduce the power consumption of networks. In the second section, we

compare three different control planes based on either centralized or distributed schemes. Moreover, we use two different slot allocation strategies (contiguous or disjoint). The performances of the proposed solutions are compared in terms of data latency, jitter, queue length and bandwidth utilization. Simulation parameters are carefully chosen to take into account implementation constraints. We find that the centralized solution with contiguous slot allocation is the most efficient as it allows a throughput up to 7 Gbps on a 10Gbps link. The computation of the burst emission patterns in the contiguous resource allocation scheme is based on a heuristic approach. However, this is done dynamically according to the variation of traffic. Alternatively, we propose another centralized allocation scheme that relies on a static resource allocation based on computing a fixed schedule taking into account only the mean throughput of the traffic; the computation is done using linear programming method. The comparison study between both alternatives shows that the static solution outperforms the dynamic one despite of the traffic variation. The work presented in this chapter was carried out within the European research project SASER. These results have been published and presented within international conferences: SoftCom 2012 [6], ICOIN 2013 [7], ONDM 2013 [8].

Chapter 5 describes MEET (Multi-hEad sub-wavElength swiTching), our proposed novel metro-backhaul network architecture. It is based on a TWIN centralized control plane. The novelty of our solution is that the metropolitan area is extended optically to reach remote nodes. Compared with current architectures, MEET proposes to aggregate traffic using passive optical nodes instead of using electrical nodes. This architecture presents a potential use-case of TWIN in an operator's network. Its architectural characteristics alleviate some constraints of the TWIN control plane. Several options regarding the dynamicity of the control plane and the burst assembly process are compared in terms of resource allocation efficiency and their robustness to the traffic variation. Performance evaluation is carried out using a simulation platform driven by real traffic traces captured on a French operator's metropolitan network. The QoS delivered to three different service classes has been assessed in terms of latency and jitter. Obtained results show that a control plane that does not adapt to short-term variations (ms range) of the real traffic may provide performance levels compatible with QoS requirements in a metropolitan network. This work was carried out in the frame of CELTIC-Plus project SASER-SaveNet, as well as the European project COMBO. It was the

subject of two papers submitted and accepted in ICTON2013 [9] and ONDM2014 [10] conferences.

In the next chapter, we propose a test-bed as a Proof of Concept (PoC) of the TWIN paradigms. The test-bed implements a static control plane for a network consisting of two sources and two destinations nodes. The control plane is developed using real-time software, while the burst emission unit is monitored by a Field programmable gate array (FPGA) software. The system provides a throughput of 10 Gbps emitting bursts of 4.5µs with a guard time of 0.5 µs. One of the main challenges of the test-bed is the difficulty of ensuring synchronization between the different components of the system. The obtained results show the excellent performance of the generated signals and their time accuracy. This work is still in progress in the frame of the SASER project; it has been used to emulate the MEET architecture. This is not presented in the present report as it was finalized at a late date.

The final chapter concludes the report, recapping the main results that have been obtained and proposing some directions for further work.

Chapitre II. State of the Art: Architectures and Protocols

Nowadays, several networks use the optical fiber as the physical medium to transport data between geographically remote sites. However, the optical fiber is just the physical support of a stack of superposed protocol layers, ensuring the end-to-end transmission. Each of the upper layers relies on one or several protocols to perform a specific task in the telecommunication process such as time division multiplexing or data routing. The protocol acts by processing data in the electrical domain and it should be able to interact with protocols of adjacent layers.

The integration of a new paradigm in the network as *Optical Burst Switching* (OBS) requires a deep study of the scenarios where it will be potentially deployed. This allows, on the one hand, identifying the adequate location on the network where this paradigm could bring the most profit and, on the other hand, determining its position in the protocol stack and the features that it should provide to cohabit with the existing protocols.

Therefore, we present, through this chapter, a general overview of the context taking into account the current and the future state of the telecommunication networks.

In the first section, we give an overview of the principal components of operators' networks. Here, we emphasize the most common technologies in the access area and the principal aggregation nodes of traffic and their positions in the metro-core network. As the architecture is often different from one operator network to another, we concentrate our description, in some parts of this section, on the network architecture of the French operator "Orange".

In the second section, we define the protocol stack of a broadband fixed network based on an optical fiber medium. Besides, we describe in details the most important protocols of this stack, stressing on the mechanisms and features that can concern OBS paradigm.

In the final section, we discuss the most likely scenarios of OBS taking into account the traffic trend. Then, we study the position that fit well to the OBS layer on the protocol stack. Here, we suggest protocols that OBS layer could substitute or interact with.

II.1. Network architecture overview

Telecommunication networks are used to connect a large group of users spread over a geographical area. The Internet Service Providers (ISP) operate a domestic IP backbone built on its own or leased from a third-party operator, which transfers all types of traffic (voice, Internet, TV, Video on Demand (VoD)) to and from a fixed/mobile residential or business users. In order to ensure an efficient connectivity, the current operator networks are designed in a hierarchical way depending on the covered area and the traffic aggregation process. A node in a given level aggregates the traffic coming from the immediate lower level, yielding to higher stages of traffic aggregation.

As shown in Figure 1, we can define three levels of hierarchy: access, backhaul and backbone. At the access level, the network covers a local area and a broadcast star is often used to combine multiple users' lines. At the backhaul level, several access networks are connected with each other. A ring topology is commonly used at this level to link backhaul nodes. At the backbone level, several backhaul networks are connected by means of a national network. At this level, nodes are generally interconnected according to a mesh topology.

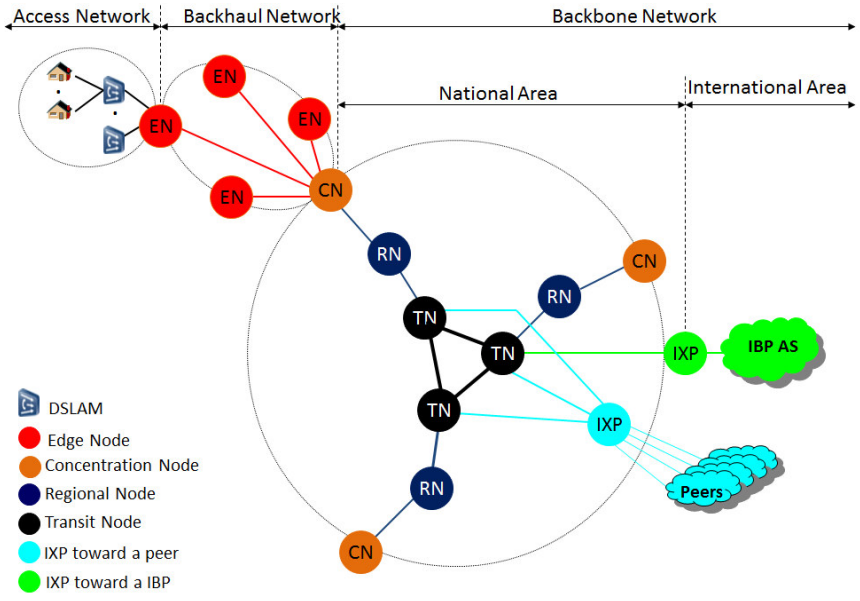


Figure 1- General view of an operator network

II.1.1. Access network

The access network is the nearest stage of the telecommunication network to the end user. It runs from the service provider to the home or business. The access networks can be fixed or mobile.

II.1.1.1. Fixed access network

The fixed access network typically consists of a hub, also called *Central Office* (CO) or *Head-End* (HE), *Remote Node* (RN) and *Network Interface Unit* (NIU) as shown in Figure 2. The CO may be connected to several RNs, with each of them in turn serving a separate set of NUIs. An NIU either may be located in a subscriber location or may itself serve several subscribers. The network between the CO and the RN is called *feeder network*, and the network between the RN and the NIUs is called *distribution network*. The role and the complexity of each element depend on the technology.

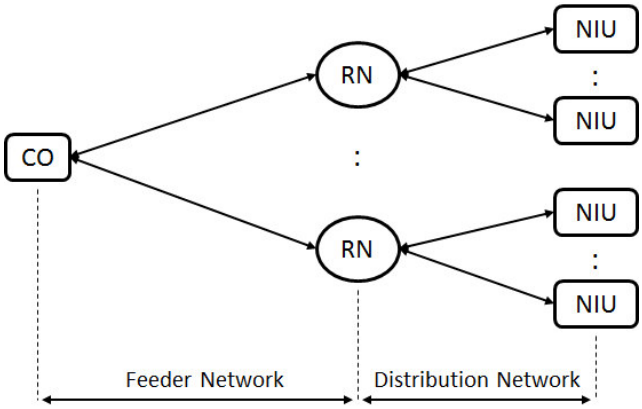


Figure 2- Typical fixed access network architecture

Previously, the fixed access network was mainly intended to provide telephone service to home. The telephone network runs over twisted pair of copper cable, which is made up of a pair of copper wires twisted together and links each customer to the CO. The telephone network was designed to originally provide 4 kHz of bandwidth to each user. Hereafter, this type of access used a modem in order to provide a narrowband access to the Internet, with a maximum download bandwidth of 56 Kbps [11].

Several approaches have been used to upgrade this access network infrastructure to support the internet and the other set of new services such as IP telephony, IP television and VoD. The fixed line technologies described here include:

- Hybrid Fiber Coax (HFC)
- Digital Subscriber Line (xDSL)
- Fiber To The x

The cable network, also called Hybrid Fiber Coax (HFC) network [12], is a broadcast network with a simple management, in which all users share a common total bandwidth. As shown in Figure 3, HFC network consists of fibers between the HE and the optical node and a coaxial cable from the optical node (analogous to RN) to the end-user (analogous to NIU). The HE delivers the same set of signals to all the end-users. The downstream channels (between the HE to the end-user) occupy a band of frequency between 50MHz and 550MHz while the upstream channels (from the end-user to the HE) occupy a band of frequency between 5 and 40MHz. The advantage of the cable network is that they are less distance limitations to extend the network than xDSL. Whereas, this technology relies on a shared network architecture, which makes the amount of bandwidth delivered to the customer dependent on how many people share the connection back to the head-end.

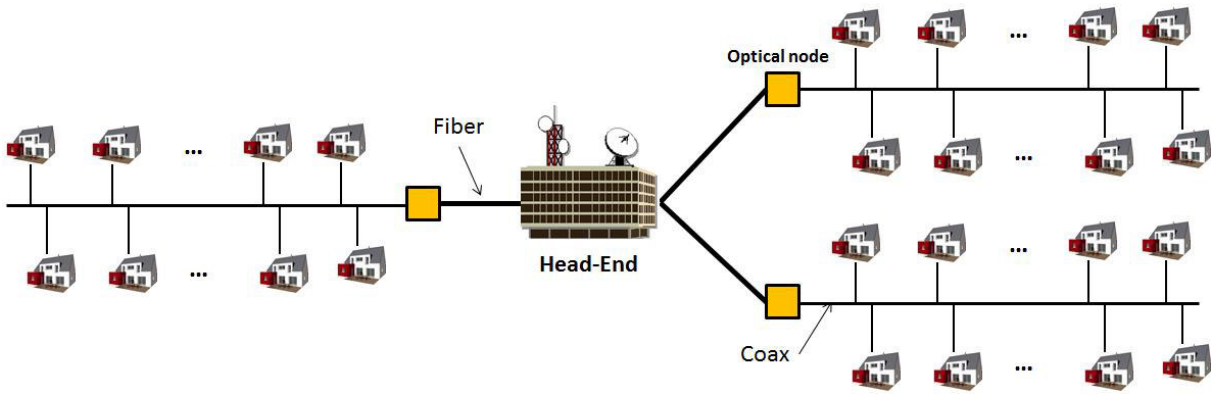


Figure 3- General architecture of HFC networks

The Digital Subscriber Line (xDSL) is a technique that works over the copper infrastructure and provides a high speed data digital transfer thanks to a sophisticated modulation and coding methods. This technique can be used for VoIP and broadband access: Internet, Multimedia, IP television, etc. However, DSL has some limitations among which are: (a) the

realizable bandwidth is strongly depends on the distance between the CO and the home, (b) the upstream path is limited to few hundreds of Kbits per second. The Asymmetric DSL (ADSL) flows are concentrated by multiplexers or Digital Subscriber Line Access Multiplexers (DSLAM), which give access to the IP network. In the case of Orange's network, an average of 900 users is connected via a copper cable to a given DSLAM [13]. Beside the ordinary users, the DSLAM can be also connected to another DSLAM via a point to point link. In this case, the first one has the role of the master and the second one has the role of the slave. The analog telephone communication and the ADSL flows circulate on the same copper pair up to the RN mainframe, occupying two frequency bands.

Fiber To The x (FTTx) provides another way to deliver access services to end users based on optical fiber. It can extend the available broadband ADSL service offer to include upstream and downstream very high bandwidth (up to 100 Mbps per user in 2011 [14]), with improved response time and reachability. Compared with ADSL, the distance between the CO and final customer is significantly larger and it has not impact on delivered bandwidth. For instance, the 10G-PON standard [14] supports a range of optical budgets from 33 dB to 35 dB. A PON with a 35 dB optical budget could span 25 km or more and be shared/split among 128 subscribers. Depending on how close the fiber gets to the subscriber, we usually distinguish a set of FTTx optical connection architectures: Fiber To The Building (FTTB), Fiber To The Home (FTTH), Fiber To The Curb (FTTC), etc... Some of these architectures are depicted in Figure 4 . In FTTx, data is transmitted digitally over optical fiber from the CO to fiber terminating node called Optical Network Units (ONUs). The RN is a simple passive device such as an optical star coupler, and it may be collocated in the CO.

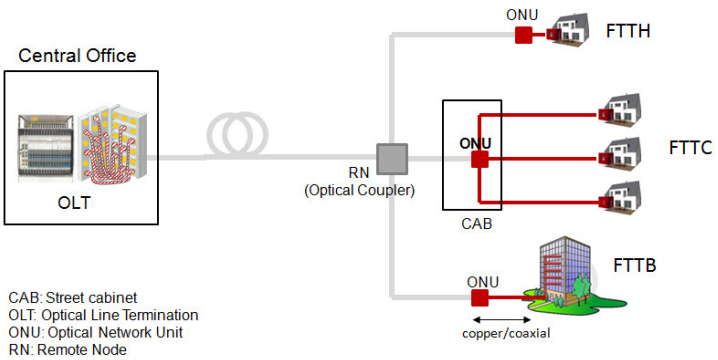


Figure 4- General architecture of FTTx networks

The network from the CO to the ONU is typically a Passive Optical Network (PON). It has a tree structure where ONU presents a leaf of the tree. The root of the tree is presented by active equipment located in the CO, called Optical Line Termination (OLT). The OLT ensures the interconnection of PON with the backhaul network and diffuses data coming from the backhaul network and service platform toward the PON. OLT consists of maximum around 128 ports. Each port is connected to one PON which can serve 64 ONUs [12]. As shown in Figure 5, the DSLAM and the OLT chassis can be collocated in the same CO. Each card manages separately its own access network.

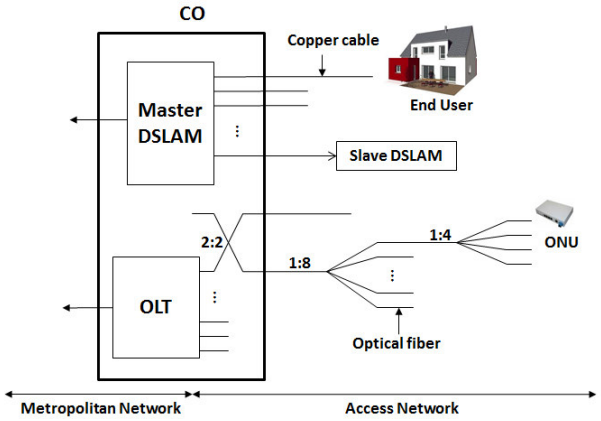


Figure 5- Collocation of DSLAM and OLT in the same CO

Optical transmission is less power consuming than electrical transmission and passive networks are not powered except in the end points, which provide significant cost saving to operators. In addition to that, the fiber infrastructure is transparent to bit rates and modulation formats, which is more accommodating to future upgrade.

In the literature, all generations of PON standards are proposed to ensure the transfer of data between end points in the downstream direction (from the OLT to the ONU) and the upstream direction (from the ONU to the OLT). These variants are based on the following principles:

- in downstream side, traffic is broadcast by a transmitter at the OLT to all the ONUs using a passive coupler,

- in upstream side, the ONUs share a channel via a multi-access protocol (e.g., Time Division Multiplexing (TDM) protocol) and an optical combiner device (e.g., a coupler).

In the TDM approach, the ONU needs to be synchronized to a common clock. This is done by a process called ranging, where an OLT measures the delay with its attached ONUs and adjusts its clock such that all ONUs are synchronized relatively to it. In some variant of PON like TPON (originally called PON for telephony), a fixed time interval is allocated to each ONU for the upstream direction. In Other variants like GPON [12], BPON [15] and EPON [16], the attribution of time interval to ONUs is based on a Dynamic Bandwidth Allocation (DBA) algorithm [13]. In this algorithm, ONUs send information about their upstream bandwidth need to the OLT. The OLT determines time intervals when each ONU can transmit upstream, and sends this information to the ONUs in the form of grants. Both ITU and IEEE PON standards (e.g. [12]) describe DBA framework (frame structure, type of messages …) without specifying exactly how to allocate bandwidth [12]. Hence, DBA algorithm is still open to suggestion [17] [18] [19].

The French operator Orange is currently deploying GPON, which is using one wavelength for the downstream (1490 nm) and one wavelength for the upstream traffic (1310 nm). The downstream bandwidth can be either 1.2Gbps or 2.5Gbps and the upstream bandwidth can be either 155 Mbps, 622Mbps, 1.2Gbps or 2.5Gbps.

II.1.1.2. Mobile access network

Many groups of standardization are working to develop the architecture of the radio access network. The main objectives of these architectures are the insurance of wide area coverage with low latency, high mobility and high data rate availability per user. This work gave birth to a number of standards and contributes to the evolution of the mobile network architecture. The evolution of the most popular mobile technologies occurred in three main steps:

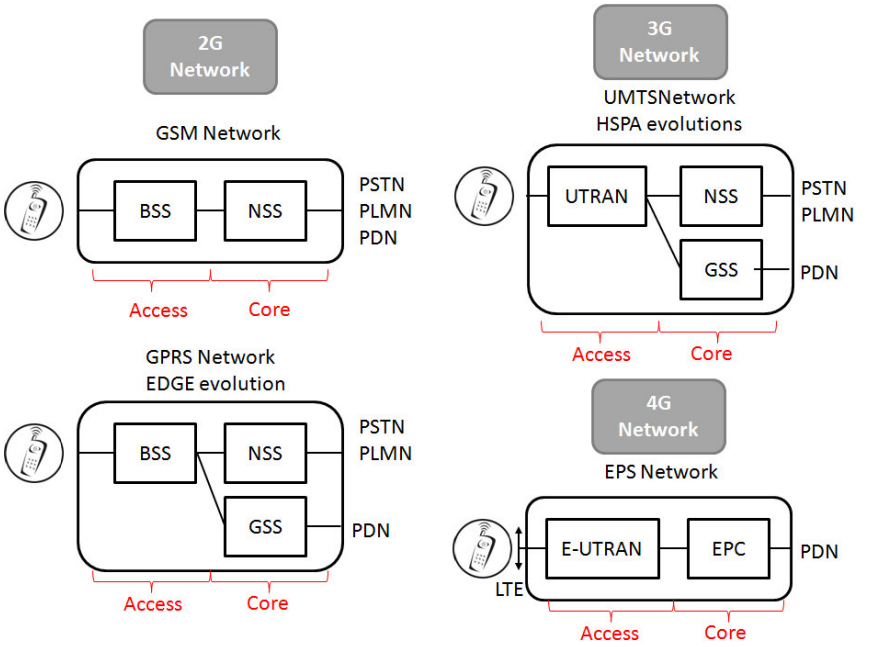
- 2G technologies: such as the Global System for Mobile (GSM) network [20], the General Packet Radio Service (GPRS) network and the Enhanced Data for Global Evolution (EDGE).

- 3G technologies: such as the Universal Mobile Telecommunication System (UMTS) technology [21] and HSPA (High Speed Packet Access) technology
- 4G technologies: such as Evolved Packet System (EPS) network [22].

The architecture of these mobile networks is composed of two subsystems: the mobile access network and the mobile core network. The mobile access network subsystem can be used to allocate the radio resource to the mobile, so that it can be either dedicated or shared. It is significantly impacted by successive evolutions. While, the mobile core network subsystem connects the access networks and one of the following third party networks:

- PLMN (Public Land Mobile Network), which is the collection of networks providing mobile telecommunications services to the public,
- PSTN (Public Switched Telephone Network), which is the collection of interconnected voice-oriented public telephone networks,
- PDN (Packet Data Network), which is the concatenation of the IP-based packet-switched networks, providing data transmission services for the public.

In the case of 4G network, the E-UTRAN (Evolved Universal Terrestrial Radio Access Network) and the EPC (Evolved Packet Core) present respectively the access and the core mobile subsystems. Unlike 2G and 3G network architectures where voice and data are processed and switched separately, 4G technology unifies the processing of the voice and the data on a unique packet-switched architecture based on Internet Protocol (IP) service. Figure 6 shows the access and the core mobile subsystems of the above-mentioned networks.



BSS : Base Station Sub-system // **NSS :** Network Switching Subsystem // **GSS :** GPRS Sub-System
UTRAN : Universal Terrestrial Radio Access Network
LTE : Long Term Evolution // **E-UTRAN :** Evolved UTRAN // **EPC :** Evolved Packet Core

Figure 6- Architecture of the mobile technologies

In one of the referenced paper [23], Cisco reported that worldwide mobile traffic, which includes Internet traffic that travels over 2G, 3G and 4G mobile access technology, will be 11 times higher in 2018 compared to 2013, reaching more than 15.9 Exabytes (EB) per month. In order to support the continuously increasing network capacity required by the mobile users, the next generation of cellular mobile phone systems will be based on smaller cell size. This huge number of cells and microcells needs to be interconnected while maintaining a low cost and a rapid response to the instantaneous variations in traffic demands. In this context, one of the promising solutions is based on transferring radio signals over optical fiber between the cell and the central station. In this case, all the complex functions will be performed in the central station while the remote antenna needs only to modulate the radio frequency subcarrier onto an optical carrier for distribution over the fiber network. This technique is known as Radio over Fiber (RoF) [24].

II.1.2. Metro-Backhaul network

The backhaul network is a metropolitan network, which represents the intermediate portion of the network between the access and the core network. It connects the access network and the

core network via the PoP (Point of Presence). It is based on a set of Edge Nodes (ENs) which aggregate flows coming from DSLAMs and OLTs. EN represents, then, the main concentration point of fixed and/or mobile traffic in the metropolitan area level. It is placed in medium sized cities inside the metro backhaul networks. In Orange network, the number of these devices is in the range of 10 to 30 in each backhaul network. Each EN aggregates in average, typically, the traffic from 64000 users [13].

Ring is the most common topology used to interconnects ENs. In the case of a large metropolitan area, regional network may be designed in the form of several interconnected rings. In this case, a primary ring which is connected to the backbone network through the PoP, collects traffic flows from several secondary rings.

In addition to these nodes, other relevant devices, such as Multiservice Nodes (MN) and Broadband Access Servers (BAS), are part of the metro-backhaul network architecture. MN provides access to the managed service platforms of the operator as VoD, TV and VoIP services, while, BAS is a broadband concentrator that aggregates the internet traffic coming from DSLAMs or OLTs and injects them in the IP network. It also sets up users' Point-to-Point (PPP) sessions.

Backhaul networks are traditionally made up of Synchronous Digital Hierarchy (SDH)/ or Asynchronous Transfer Mode (ATM) [25] technologies for mobile and fixed aggregation. Other technologies are rolled out, such as Ethernet technologies and IP/MPLS technologies to replace the ATM technology. These protocols will be further described in section II.2.

Researches intended to metropolitan networks are still active and new solutions based on optical switching are proposed. The main goal of those proposals is to cope with the increasing volume of data sent over fixed and mobile networks and, at the same time, to optimize costs. Some of these solutions will be detailed in the next chapter (Section III.2).

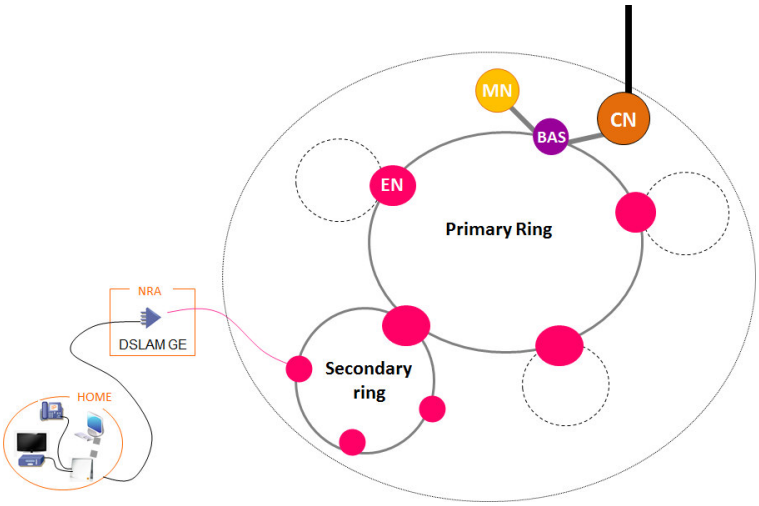


Figure 7- Simple model of a backhaul network

II.1.3. Backbone network

The core network is a national network interconnecting the backhaul networks and providing them access to the international networks. It refers to the backbone of the telecommunication network and is built over very high bitrates transmission links, connecting the principal nodes of the network. Each service provider designs its network architecture taking into account some factors as the covered area, the traffic load, the geographical characteristics, etc.

In the French national network, the Concentration Node (CN) stands for the gateway from the backhaul networks to the backbone network. The traffic coming from all the ENs of a metro area is concentrated in the corresponding CN. These nodes represent, then, the first element of aggregation in the core network. Placed in large cities, they are as many as metro networks. A CN is generally attached to a Regional Node (RN). RN aggregates traffic coming from a set of CN. It is then the second stage of traffic grooming in the core network. It is directly attached to the Transit Node (TN), which provides the interconnection between the national network and the international networks.

This historical architecture of the core network is hierarchical and centralized around the TNs. As the amount of traffic is increasing and in order to alleviate the aggregation load in the TN, this architecture undergoes some modifications. Actually, additional nodes, called Internet Nodes (IN), are created in the backbone area ensuring the connection of the CN to the global internet network. Moreover, in some cases, CN are connected directly to the TNs without going through the RN and RN can be, in some cases, a gateway to connect to another operator

network. These modifications enable the backbone network to move toward a more distributed architecture.

The largest national network of the French operator "Orange" is located in France and known as the RBCI (Réseau Backbone de Collecte IP). It deserves almost ten millions of Internet subscribers. The RBCI is an IP *Autonomous System* (AS3215).

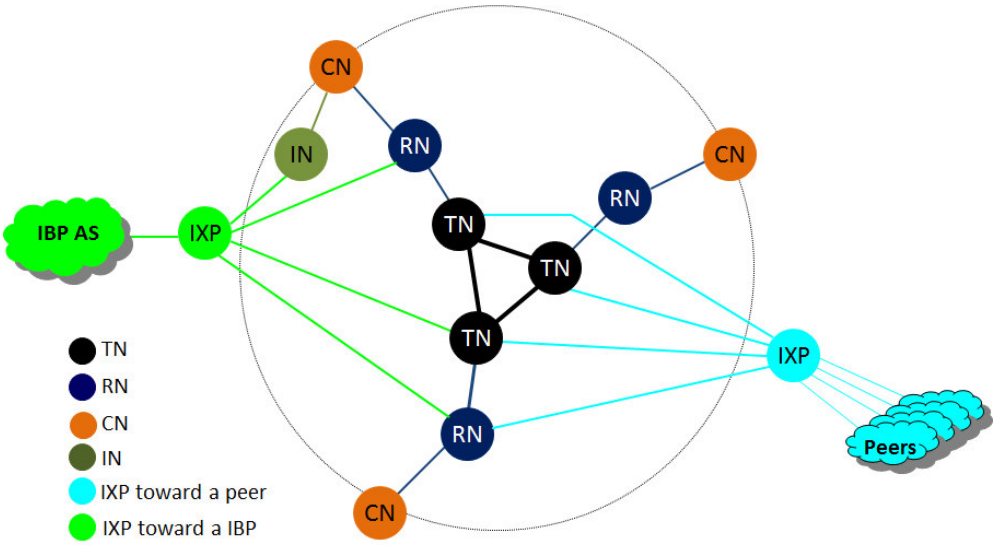


Figure 8- Simple model of a backbone network

The AS represents the fundamental granularity to describe the global Internet. Two major kinds of AS can be identified: (a) the ISP (Internet Service Provider)' autonomous system, which offers Internet access to residential and/or business customers and (b) the IBP (Internet Backbone Provider) autonomous system or transit network, which offers transit to ISPs in order to ensure global Internet connectivity. ASs exchange traffic via a physical infrastructure called IP *interconnection points* (IXP) and they are interconnected in various manners depending on their respective sizes and geographical spans. In order to offer transit to its customers, an ISP has to choose either to peer with other ISP's autonomous systems, or to rely on transit offered by IBPs. In the former case, the different ISPs exchange internet traffic between their networks (AS) by means of mutual peering agreements, which allow traffic to be exchanged without cost. In the latter case, ISPs have to pay IBPs for transit in order to provide full Internet connectivity to their customers [26].

Some ISPs have their own IBP. This is the case of the French operator Orange which exchanges its traffic with the rest of the world via its international IP network, known as Open

Transit Internet (OTI). It aims to provide global Internet connectivity to the group subsidiaries', operator customers', ISPs' and content providers' IP networks.

II.2. Protocols in telecommunication network

In this part, we focus on the possible protocol stacks that we can find in a WDM-based transport network. Then we describe the most common protocols over a WDM optical network. The goal of this study is to identify the position of OBS layer(s) and the characteristics of its adjacent layers. Moreover, this description helps to understand mechanisms which are used in designing some OBS solutions and are inspired from existing protocols.

II.2.1. Protocol stack overview

In the past, the carrier networks were designed to support connected traffic, and the data traffic was transmitted using the voice channels. Now, the core networks are being designed for supporting packet traffic. In addition to enhancement of services and network capabilities, new protocols have appeared and other ones are upgraded in order to satisfy the rapidly growing demands for bandwidth and the need for better quality of service (QoS), protection, availability, etc.

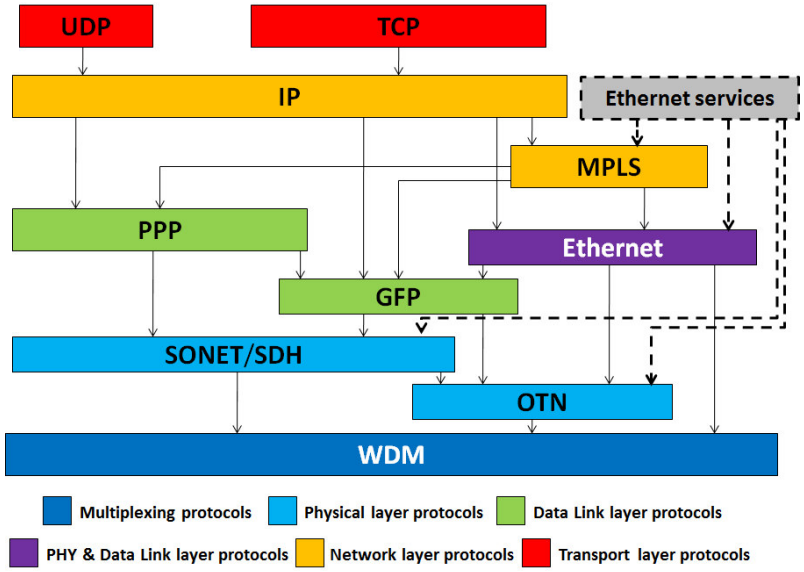


Figure 9- Protocol stack of a telecommunication network

Wavelength Division Multiplexing (WDM) system is the basic technology, on which network operators rely on to offer wide bandwidth on optical fibers and huge transmission capacity in the core network. WDM significantly increases the fiber capacity utilization by dividing the available bandwidth into multiple wavelength channels. Wavelengths are modulated separately and sent into the fiber simultaneously. As long as the power within each signal is not too high, the fiber acts as a linear medium, the interaction of different wavelengths on each other will be negligible, and each wavelength propagates in the fiber independent of the others.

Protocol stack of the core network is compliant to the OSI model, where communication system is partitioned into protocol layers. Each layer benefits from service provided by the below layer and executes a specific task serving the layer above it. Figure 9 presents an attempt to find the possible interaction between protocols in a core optical network and to classify these protocols according to the OSI model.

The optical layer provides the physical link to the upper layers which process the data in the electrical domain (such as fixed time division multiplexing or aggregating a variety of bit-rate services into the network). The upper physical layer can operate over point-to-point fiber links as well as over a more sophisticated optical layer, using an all-optical channel established between the end-nodes, namely a *lightpath*. The predominant physical layer protocol (according to the OSI hierarchy) in the core networks today are Synchronous Optical NETwork/Synchronous Digital Hierarchy (SONET/SDH) [1], Ethernet [27], and the Optical Transport network (OTN) [28].

SONET/SDH as part of the first generation of optical networks was the earliest to be deployed in backbone networks and has been very successful over the years. It has been designed for Constant Bit Rate (CBR) connections, and it can add and drop CBR flows (called Virtual Containers) on a Synchronous Transport Module (STM) at different line rates (155M to 10G, typically) by using time division multiplexing. It can transport packets thanks to data link layer protocols that adapt packets to SDH containers (maximum size VC4 of 150Mb/s). In order to map the native traffic to the SONET/SDH containers, an adaptation mechanism such as Generic Framing Procedure (GFP) [29] is used. GFP works for a variety of data protocols, including IP, Ethernet and Multi-Protocol Label Switching (MPLS). It is

used to adapt native data traffic to an incumbent transport network infrastructure and to provide an efficient and QoS-aware mechanism to map packet data to a CBR channel. GFP is particularly suitable for SONET/SDH, OTN links, or even for dark fiber applications.

OTN is built upon some concepts of SONET/SDH and has been designed to carry all types of data traffic including SONET/SDH and Ethernet traffic. It has been designed to operate from tributaries at 1G to very high transmission line rates (2.5 G to 100G, available), and it has a complete and flexible set of operation and management features. This protocol will be described in more details in the next section (II.2.2.1).

Ethernet is carried over all communication media including coaxial cable, twisted pair, wireless, and fiber optic cables; in addition to that, it can be carried over other physical layer protocol infrastructures already installed by operators. The most common Carrier Ethernet Transport (CET) methods in optical network are Ethernet over SONET/SDH, Ethernet over OTN, Ethernet over ATM... Mapping can be done bit to bit directly or a link adaptation protocol as GFP can be used. In this case, only the useful data of Ethernet frames are transported, extra coding bytes are discarded. A well-known interface is the 10 Gegabit Ethernet: it fits into SDH or OTN containers under the Wide Area Network (WAN) implementation (9,95G) but not under the Local Area Network (LAN) one (10.31G). This issue has been “solved” in the case of OTN, by allowing different line transmission rates for OTU2 and OTU3.

As per OSI concepts, Internet Protocol (IP) and Multi-Protocol Label Switching (MPLS) are not proper upper layers of the optical layer. An underlying physical and data link layers are required to ensure their transport over the optical paths. Most carrier networks employ an overlay model migrating toward a simplified architecture basically composed of four layers: MPLS (layer 3), GFP (layer 2), OTN (layer 1) and WDM (layer 0) [30].

Some studies propose alternative architectures as *IP over WDM* [31] [32] [33], where IP is integrated closely to the WDM optical layer. As processing of IP packets in the photonic domain is unfeasible in the foreseeable future because of the lack of photonic memories, MPLS is used as an integration structure between IP and the underlying layer. This

architecture, referred to as *Multi-Protocol Lambda Switching* (MPλS) [34] [35], represents an extension of MPLS concept to provision light circuit.

II.2.2. Description of some protocols

II.2.2.1. Optical Transport Network (OTN)

The Optical Transport Network (OTN) [28] was designed to extend capacity transport of SDH and to better cope with data packet traffic such as IP and Ethernet, as well as the previous transport technology in particular SONET/SDH. It was created with the intention of combining some benefits of SONET/SDH technology (OAM mainly), the integration of WDM channels management and bandwidth expansion capabilities (creation of higher rate container (OPUx) than SDH ones (VCx)). Note that OTN is an asynchronous technology: to drop one tributary, all the OTUx must be demultiplexed. In general, the OTN consists of three optical layers (Optical Transport Section (OTS), Optical Multiplex Section (OMS), Optical Channel (OCh)) and three digital layers (Optical Transport Unit (OTU), Optical Data Unit (ODU), Optical Channel Payload Unit (OPU)). These layers are depicted in the Figure 10 below [36]:

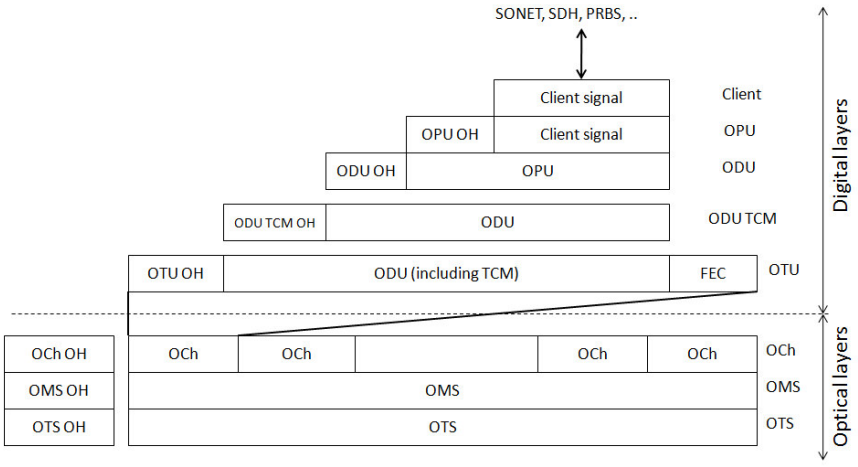


Figure 10- OTN layers

The OTU encapsulates two layers: ODU and OPU, which provide access to the payload (SONET, SDH, etc …) and it standardizes Forward Error Correction (FEC) mapping for the WDM channels. It allows an increase in the optical link budget by providing a method to correct errors, thereby reducing the impact of network noise and other optical phenomena

experienced by the client signal traveling through the network. As a consequence, FEC allows operators to increase the range of the sections between regenerators, and thus to lower costs.

To create an OTU frame, a client signal rate is first adapted at the OPU layer. The adaptation consists of matching the client signal rate to the OPU rate, sometimes by stuffing. Once adapted, the OPU is mapped into the ODU. The ODU also adds the overhead necessary to ensure end-to-end supervision and Tandem Connection Monitoring (TCM). Finally, the ODU is mapped into an OTU, which provides framing as well as section monitoring and FEC.

As shown in the Figure 11 below, the OTU frame is broken down into the following components: frame alignment overhead, OTU/ODU/OPU overheads, OPU payload and OTU FEC.

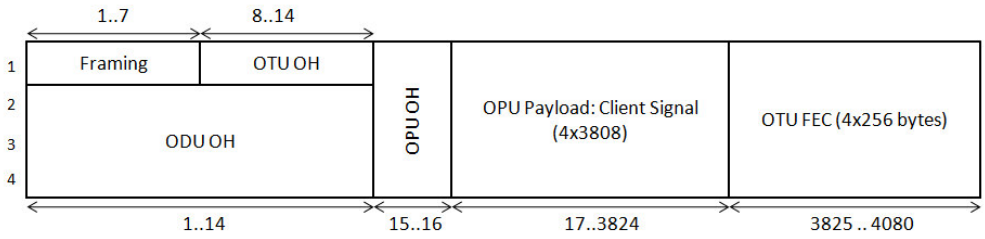


Figure 11- OTU frame structure

OTN could be bit-transparent. An operator can offer services at various bit rates (2.5G, 10G …) independent of the bit rate per wavelength using the multiplexing and inverse multiplexing features of the OTN [28]. It maintains the integrity of the whole client signal. It could be also timing transparent, as the asynchronous mapping mode can transfer the input timing to the far end [28].

II.2.2.2. Ethernet

Ethernet was originally designed for simple data sharing over a LAN in campuses or enterprises. But now, it is spreading in the next-generation carrier networks. The line rate and transmission range of Ethernet networks is steadily increasing, by a factor 10 every release. Many standards have sprung from in order to develop new services, in particular carrier

networking. Ethernet is no longer just a shared access medium, which allows avoiding collision, but it is also a data link layer and physical layer technology.

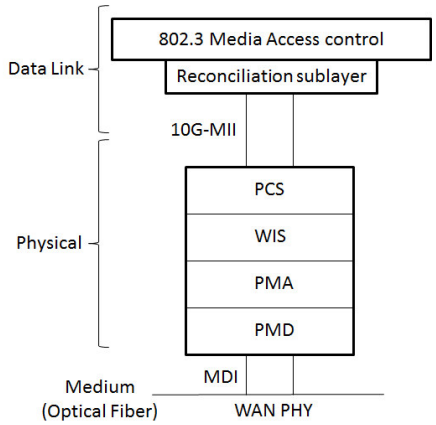


Figure 12- 10 Gigabit Ethernet protocol

The physical (PHY) layer converts the data coming from the Medium Access Control (MAC) layer into optical or electrical signals and sends it across the physical transmission medium. 10-Gigabit Ethernet links can be used to connect LAN traffic to WAN. In such scenario, end user or application traffic can be aggregated through 1-Gigabit Ethernet and then 10-Gigabit Ethernet and connected to WAN for long distance transmission. WAN PHY operates at a rate compatible with the payload rate of STM64 (9.62 Gbps) to provide support for transmission of Ethernet on networks based on SDH. As depicted in the Figure 12, Media Independent Interface (MII) is the interface between the MAC layer and the physical layer. It allows the same MAC layer to connect various media types.

Ethernet has a MAC protocol to arbitrate transmission between nodes. The most famous arbitration protocol, referred to as Carrier Sense Multiple Access with Collision Detection (CSMA/CD) [37], was originally intended to the local networks. According to this protocol, if a node has a packet to transmit, it should listen to the link. When it detects that the link is idle, it transmits its packet and at the same time, it continues listening. If it detects a collision, then it stops packet transmission and it waits during a randomly chosen delay before reattempt the transmission. The collision detection algorithm of the CSMA/CD mandates that round-trip propagation delay between any pair of stations must not exceed the transmission time of the smallest data frame. Hence, acceleration of Ethernet to Gigabit speeds has created some

challenges. In order to increase the diameter of Gigabit Ethernet network, a *carrier extension* [38] has been added to the Ethernet specification. This process adds bits to the frame until the frame meets the minimum slot-time required. The minimum frame size is extended from 512 bits to 512 bytes. However, carrier extension decreases the bandwidth efficiency for small frames. To overcome this problem, another change to the Ethernet specification is proposed: *frame bursting* [38]. Frame bursting is an optional feature in which an end station, in a CSMA/CD environment, can transmit a burst of frames over the wire without having to relinquish control. Other stations on the wire defer to the burst transmission as long as there is no idle time on the wire. The transmitting station that is bursting onto the wire fills the inter-frame interval with extension bits such that the wire never appears free to any other end station.

The structure of a basic Ethernet frame is shown in Figure 13. Destination and source addresses identify the receiving and sending station for the frame. The length/type field represents the number of valid data octets contained within the data field. The data field carriers the payload information. The "pad" field is used to fill out the frame to the minimal size i.e. 64 bytes, necessary for collision detection. The last four bytes corresponds to the Frame Check Sequence (FCS). It encodes a checksum based on the frame contents excluding the first eight octets.

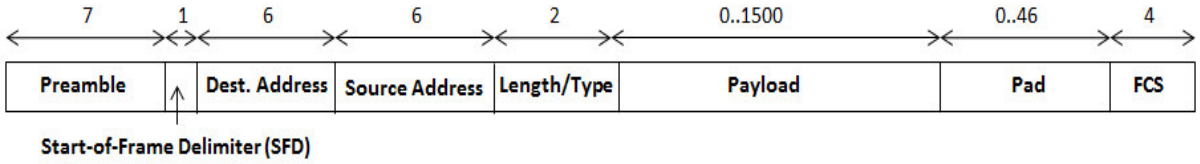


Figure 13- Ethernet frame structure

MAC layer of Gigabit Ethernet supports both full-duplex and half-duplex transmission [38]. For half-duplex transmission (shared access), CSMA/CD is utilized to ensure that stations can communicate over a single wire and that collision recovery can take place, whereas the full-duplex provides the means of transmitting and receiving simultaneously on a single wire. Full-duplex has allowed bandwidth on Ethernet and Fast Ethernet networks to be easily and cost-effectively doubled. Since the end nodes do not interfere with each other's transmission, CSMA/CD becomes unnecessary. In this case another link-level flow control

mechanism called *pause mechanism* [39] is performed in order to avoid congestion in the receiving station.

The several versions of Ethernet transform it from a CSMA/CD based technology intended to the local area and providing low throughput to a full duplex link able to reach a throughput superior to 40G/100G and intended to the metropolitan area. The success of Ethernet and the strong demand to deploy it in the transport networks are related to many factors such as cost effectiveness, flexibility and ease of interoperability. To reach high bitrate and long distance, the evolution of Ethernet tends towards the introduction of Ethernet tunnels. The Provider Backbone Transport (PBT) is an Ethernet technology addressed to the transport network. It creates point-to-point tunnels Ethernet to provide QoS, fault resilience and OAM (Operations, Administration and Management) to the network, with a possibility of traffic engineering. It is based on Ethernet standards IEEE 802.1Q [40], IEEE 802.1ad [41], and IEEE 802.1 ah [42]. All these standards, before the definition of PBT, aimed at addressing the problem of lack of hierarchy in Ethernet.

The concept of VLAN (Virtual LAN) was introduced by the IEEE 802.1Q standard, which provides for the first time, a hierarchy in Ethernet. The *Virtual LAN* (VLAN) [40]. It allows the network bandwidth to be shared among groups of nodes, so that each group can communicate over its own VLAN. VLAN technology can be used to implement *Virtual Private Networks* (VPNs) [43] [44]. A unique VLAN can ensure the connection of an enterprise having many sites at different locations. In this case, the connection is done through a service provider which should be able to offer the Carrier Ethernet Service (CES) [45]. These services include E-LINE, E-LAN and E-TREE. E-LINE service provides a dedicated Ethernet point-to-point connection between any two points on the network. E-LAN service provides a multipoint connection that operates as a virtual switched Ethernet network. It permits multiple locations to exchange data with each other as if they are connected directly to the same LAN segment. Finally, E-TREE service provides an Ethernet point-to-multipoint connection. The Ethernet VLAN (802.1Q) frame contains 4 bytes field called Q-tag added to the basic Ethernet frame header to identify the VLAN (12bits) and the CoS. It is inserted between the source address and length/type field as depicted in Figure 14. The IEEE 802.1ad (also known as Q-in-Q) provides the customers the ability to organize the multiple VLANs,

within the service provider's VLAN. Another field Q-tag is added to support this new type of address. The IEEE 802.1ah (also known as MAC-in-MAC) provides to Ethernet true scalability of carrier-grade networks. As shown in Figure 14, the MAC client packet is encapsulated (without or with the FCS field) in the MAC service provider packet. A new service tag field of 24 bits was introduced (I-SID, Service Instance IDentifier), allowing the total distinction between customer and provider domains.

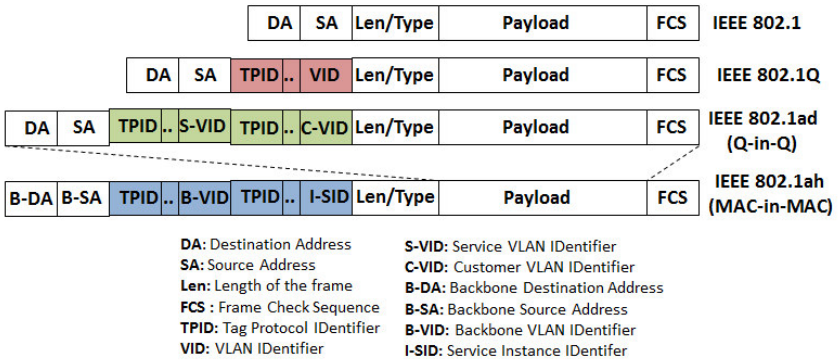


Figure 14- Frame structure of IEEE802.1, IEEE802.1Q, IEEE802.1ad and IEEE802.1ah

Many technologies can be used to carry Ethernet service such as SONET/SDH, MPLS and Ethernet itself. The characteristics of the support network may limit more or less the size and the performance of the network, or the quality of service of the borne applications. Among the drawbacks of using SONET/SDH technology as carrier of Ethernet service is the need of an adaptation layer as GFP. MPLS can be also used to transport Ethernet thanks to its *pseudowire* [46] technology. Although SONET/SDH and MPLS can already provide the service carrier features, enhancing Ethernet OAM and traffic engineering may lead to a serious concurrent. Firstly, Ethernet has traditionally been less expensive than SONET/SDH and MPLS for local networks. Secondly, it may be simpler to operate and manage a network with one protocol than a mix of protocols.

II.2.2.3. Internet Protocol (IP)

The Internet Protocol (IP) [47] is a network layer (layer 3) protocol. It transports information in form of packets, which are of variable length. IP Router forwards packets from an incoming link onto an outgoing link using addressing and control information maintained in the routing table to determine the route to the destination host. IP routers have a "network view", they are able to re-route data in case of congestion or failures. A routing protocol

(OSPF, IS-IS ...) is used by routers to ensure the delivery of packets from the source to the destination.

Role of IP was traditionally to provide connectionless and “best-effort” delivery of packets through an interconnected network. It performs fragmentation and reassembly of packet to support data links with different Maximum Transmission Unit (MTU) sizes. Best-effort service means that IP tries its best to forward a packet from its source to its destination, without regards to transmission parameters. Different packets may take different routes through the network and experience random delays, and some packets may be dropped if there is congestion in the network. There has been a great deal of effort to improve that so as to offer some Quality-of-Service (QoS) assurance to the users of the network. Within IP, a mechanism called DiffServ (Differentiated Services) [48] [49] has been proposed. In DiffServ, packets are grouped into different classes according to the type indicated in the IP header. The class type specifies how packets are treated within each router. Packets marked as expedited forwarding (EF) are handled in a separate queue and routed through as quickly as possible. Several additional priority levels of assured forwarding (AF) are also specified; an AF has two attributes: xy. The attribute x typically indicates the queue to which the packet is held in the router prior to switching. The attribute y indicates the drop preference for the packets. While Diff-Serv attempts to tackle the QoS issue, it does not provide any end-to-end method to guarantee QoS. For example, it is not possible to determine a priori if sufficient bandwidth is available in the network to handle a new traffic stream with real-time delay requirements. This is one of the benefits of Multi-Protocol Label Switching (MPLS) that will be described in section II.2.2.4.

Several layering structures are possible to map IP into the optical layer. The term IP over WDM can refer to a variety of possible mappings, having in mind to simplify this mapping.

As IP packets could be dropped across the network, protocols of the transport layer (layer 4), as Transmission Control Protocol (TCP) [50], can be used as a highly reliable host-to-host protocols. TCP provides many services such as stream data transfer, reliability, efficient flow control and full-duplex operation. Another commonly used transport protocol for simple message transfers over IP is the User Datagram Protocol (UDP) [51]. UDP, which is a connectionless transport-layer protocol, is used by many applications as Network File System

(NFS) and Simple Network Management Protocol (SNMP). UDP is used associated to RTP for real time application such as VoIP.

The internet is a global network, and it is impossible to expect each router to maintain a topology of the entire Internet. For this purpose, the network is divided into multiple interconnected domains; each domain is called an autonomous system (AS) [52]. Separated inter-domain routing protocols, such as Border Gateway Protocol (BGP) [52], are used to exchange routing between domains in a large network.

Due to the lack of IPv4 addresses, there is a global push towards IPv6 that includes bigger address space [53] [54]. According to Cisco study about traffic [23], fixed and mobile network operators globally are deploying IPv6. A notable IPv6 traffic generation, ranging from several percent of traffic to upward of 10 percent, is starting to be seen. The forecast estimates that IPv6 fixed traffic would reach 24.8 exabytes per month or 23 percent of total fixed traffic in 2017.

II.2.2.4. Multiprotocol Label Switching (MPLS)

Multi-Protocol Label Switching (MPLS) [55] is a data forwarding technology for use in packet networks, designed at the beginning to simplify ATM packet routing equipment and quickly applied to IP. But it provides more than that, as options for traffic engineering. It provides a very high-speed data forwarding between nodes together with reservation of bandwidth for traffic flows and insurance of QoS requirements. It is designed to carry data packet using established based paths. Each path is associated to an arbitrarily assigned label.

The MPLS header is called *shim header*. It is a 32-bit field inserted before a packet (could be an ATM/IP packet, an Ethernet frame and so on...) (see Figure 15) : note that MPLS transported entities could range from layer 1 to 7, but that MPLS itself needs to be transported in layer 2 frames (Ethernet, GFP, PPP...). The shim header determines the Time To Live (TTL) of the transported packet, its COS, the path that it must follow, etc.

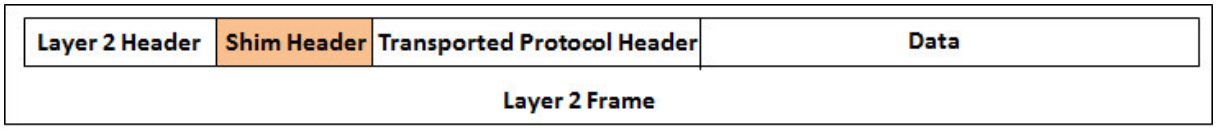


Figure 15- The position of the shim header

Each MPLS node, called Label Switching Router (LSR), determines the next hop for the packet using a look up table called Label Forwarding Information Base (LFIB) which contains a mapping of {incoming interface, incoming label} to {outgoing interface, outgoing label}. Thus, the intermediate LSRs are not obliged to examine the IP header in each hop during forwarding. Instead, they forward labeled IP packets according to the label swapping paradigm. The virtual connection that a packet follows across the network is called Label Switched Path (LSP). It is set up, modified, rerouted, and torn down by an edge router which is referred to as Label Edge Router (LER). In this context, the configuration of LFIB on each LSR and the exchange of label mapping information within the control plane between the LSRs is a complex process in a large network. To cope with these issues, signaling and routing protocols are proposed to enable MPLS to support the reservation of network resources as well as the possibility of performing constraint-based routing needed for Traffic Engineering (TE) and fast reroute (FRR). Signaling protocols are used to exchange messages within the control plane in order to establish, modify and terminate LSPs. RSVP-TE and CR-LDP are the two most known signaling protocols in MPLS networks. Whereas, routing protocols, such as OSPF, have the task of distributing information that will be used as the basis of the path computation in order to determine how LSPs will be placed within the network. Hence, the suite of MPLS protocols comprises traditional IP routing protocols (e.g., Open Shortest Path First (OSPF)) and extensions to existing signaling protocols (e.g., Resource reSerVation Protocol (RSVP)).

MPLS-TP [56] is a transport profile of MPLS. It is a connection oriented technology defined for next generation converged packet transport networks. It supports large variety of services thus it needs to be client and physical layer agnostic. The key roles defined in this technology are the implementation of OAM and resiliency features to ensure the capabilities needed for carrier-grade transport network.

Generalized MPLS (GMPLS) [57] [58] extends MPLS to provide the control plane (signaling and routing) for devices that switch packets, time slots, wavelengths, wavebands and fibers. This control plane aims to simplify network operation and management. It manages the connection provision, network resource and the QoS level.

II.3. Summary

Through this chapter, we have presented an overview of operators' networks, highlighting their main evolution toward optical fiber-based medium. Actually, the deployed network architectures and protocols are the result of a stepwise accumulation of improvements and developments aiming to confront the telecommunication digital boom challenges and to meet end-user needs.

Although the changes occurred in the network architecture, its hierarchical structure consisting mainly of three levels: access, backhaul and backbone, is still apparent. In each level of the network, traffic is aggregated and then transmitted to the upper level. Especially, traffic within metro-backhaul networks is typically an aggregation of traffic flows coming or destined to the access network. According to a study done by Cisco [23], the global average fixed broadband speed continues to grow and will nearly quadruple from 2012 to 2017, from 11.3 Mbps to 39 Mbps. Factors influence the fixed broadband speed are related to the deployment of high-speed access technologies, including the adoption of FTTH, high-speed DSL and cable broadband. For the average mobile network connection speed, it will grow from 526 kbps in 2012 to 3.9 Mbps in 2017. This high growth is due in part to the deployment of 4G where user device can exchange data at a speed up to 100Mbps.

The increasing demand of bandwidth and the development of access network, that will be more IP-oriented, impact the traffic pattern. The forecast shown in [23], global IP traffic in 2012 stands at 43.6 exabytes per month and will grow threefold by 2017, to reach 120.6 exabytes per month.

The traffic trend is not only driven by the technological progress but also by the user behavior. Indeed, according to the same study, the busy-hour traffic continues to grow more rapidly than the average traffic. In 2012, busy-hour Internet traffic grew 41 percent, while average traffic settled into a steady growth pattern. The growing gap between peak and average traffic is mainly due to the continue dominance of video traffic in all consumer internet traffic (69 percent in 2017, up from 57 percent in 2012). Video traffic has a particular consumption pattern tends to have a "prime time" contrary to the other forms of traffic that are spread evenly throughout the day (such as web browsing and file sharing).

The importance of the video traffic explains the increasingly significant role of Content Delivery Networks (CDN), which avoid bypassing backbone links to get multiple copies of the content to multiple users in the same metro-backhaul area. The CDN traffic will deliver almost two-thirds of all Internet video traffic by 2017.

As a result of this grown concentration of content sources within the backhaul network, traffic flows estimated to undergo significant change driven largely by IP. In this context, Cisco report [59] shows that in 2012, total metro-backhaul traffic was 1.8 times higher than backbone traffic, and by 2017, it will be 2.4 times higher than backbone. Another report published by Bell Labs mention that, by 2017, 75% of total metro-backhaul traffic will be terminated within the backhaul network and 25% of traffic will traverse the backbone network.

Consequently, the backbone network links remains highly loaded with a slight traffic variation. However, the backhaul networks will need more bandwidth and sophisticated resource management to cope with the dynamic variation of the traffic. These trends will have an impact on how service providers have to evolve and architect their metro-backhaul networks. Hence, they have to look for innovative and cost-effective solutions that enable agile, scalable and efficient transport of data. These solutions should not only focus on enhancing protocol mechanisms but also provide new optical layer replacing the current one based on optical circuit switching.

The Optical Burst Switching (OBS) solutions could be a good candidate in the sense that it can provide flexible mechanisms to share bandwidth and managed distributed network architectures. From a protocol point of view, three scenarios can be considered for OBS framework:

- Replace IP over MPLS over Ethernet over OTN/SDH over WDM with OBS over WDM. Performing functions of layer 3 by OBS is a very ambitious scenario and seems unrealistic for the moment since the immaturity of optical memories and inexistence of all-optical signal processing.
- Replace the Ethernet over OTN/SDH over WDM by OBS over WDM. This scenario is feasible if the OBS successes to perform efficient burst assembly for Ethernet

frames and meets the performance in terms of delay, throughput and Packet Loss Ratio (PLR).

- Replace OTN/SDH over WDM by OBS over WDM. This is the minimum level that OBS can occupy. OBS is then seen as a transport layer: lightpaths have to be established in advanced and no switching task is attributed to OBS. This scenario under-utilizes the switching capabilities envisaged for OBS and doesn't benefice from the all-optical switching mechanisms offered by OBS.

Chapitre III. **State of the Art: Optical Switching Solutions**

The increasing traffic volume everywhere in the network motivates service providers to increase bit rates at different network levels (access, metropolitan and core). 100 Gbps is now the common bit rate for long haul transmission links and the question of efficiently filling these big pipes is a real issue.

Nowadays, an entire wavelength bandwidth is reserved to ensure connection between each couple of nodes in the network. Nevertheless, network nodes may request connections at rates that are lower than the full wavelength capacity and then this per-wavelength granularity reservation could offer huge bandwidth that surpasses the real connection's needs. To minimize this bandwidth wastage, network nodes aggregate flows in order to transport them in the same wavelength. Here, we mean by flow the stream of traffic transmitted from a first node, called source node, to a second node, called destination node, within the same network. The aggregation optimizes the optical resource utilization but till now it can be performed only in the electronic domain since the non-existence of optical data processing technology. Indeed, the aggregated traffic has to be converted from the optical domain to the electronic domain in order to be processed before being converted back into the optical domain. This complex process consumes significant electrical power and generates extra latency.

In this context, the sub-wavelength switching was proposed as an alternative paradigm for this traditional wavelength switching. It aims to share the same lambda between many flows of traffic without resorting to electrical aggregation. The optical aggregation is then performed by transparently switching traffic at a granularity finer than the wavelength.

The optical aggregation could be done by multiplexing data in the frequency domain (frequency-domain sub-wavelength) or in the time domain (time-domain sub-wavelength).

The frequency-domain sub-wavelength switching [60] is based on the division of the WDM channel spectrum into several independent sub-bands that are used to transmit a low-rate flow. Flows having a common segment or path can be grouped into the same wavelength. Then, sub-bands are switched independently whilst remaining in the optical domain.

In the other side, time-domain sub-wavelength switching divides the wavelength into slices of time enabling the transmission of data into suitably sized blocks that could be packets or bursts of packets. Thus, each node shares the same interface to communicate with other nodes of the network. The same transmitter can be used to forward traffic to different destination nodes, and by the same way, the same receiver can be used to receive traffic from different source nodes. In this way, a transmitter of a node and a receiver of another given node are not constrained to communicate only one with the other, but they can be freely matched with other network interfaces according to the current needs. In this chapter we will only focus on time-domain sub-wavelength switching and to simplify terminology, we will simply refer to it as sub-wavelength switching.

We note here that some sub-wavelength switching solutions like Optical Packet Switching (OPS) [2] will not been addressed since they currently seem far from being deployed. Indeed, these solutions require specific technological components that are not available at the moment such as optical memories.

This chapter is organized as follows. In the first section, we describe the optical switching solutions, currently deployed in the operational networks. Then, in the second section, we describe the proposed sub-wavelength switching solutions based on Optical Burst Switching (OBS) paradigm.

III.1. Optical circuit switching solutions

Nowadays, transport networks rely on Optical Circuit Switching (OCS) technologies. Since the appearance of WDM technique different circuit switching paradigms have been introduced during the years.

III.1.1. Opaque switching

In the opaque circuit switching, the optical channel is converted into the electrical domain as it passes through the node. Indeed, when a wavelength is detected at a node, an optical-to-electrical conversion is performed. The traffic is then processed in order to drop or forward some flows. In the case of forwarding, the electrical signal is converted back into optical domain (electrical-to-optical conversion) and sent into fibers towards its destination. This process is referred as optical-electrical-optical (O-E-O) conversion.

The role of optics in these networks is limited mainly to data transmission. Each node has access to the signals in the electrical domain and can therefore perform extensive performance monitoring (signal identification and bit error rate measurements). The bit error rate measurement can also be used to trigger protection switching. Furthermore, the intermediate node can provide wavelength conversion, signal regeneration and low-speed grooming. Moreover, it can exchange information with other network elements by using in-band overhead channels embedded in the data stream.

However, the O-E-O process is costly and generates system complexity. In fact, the electrical switch cores require separate port cards for each network interface to convert the input signal into a format suitable for the switch fabric. Moreover, this process is very energy consuming and the large increase of the traffic volume has made the energy requirements even larger.

Thus, it would be useful to find ways to keep the signal within the optical domain, and only convert it to the electrical domain at the destination in order to overcome the heavy electronic processing load.

III.1.2. Transparent switching

In the opaque configuration, the optical signal is converted into the electrical domain as it passes through an intermediate node along its path. However, in transparent configuration [61], the optical channel, or wavelength, between two network nodes is optically switched at the intermediate nodes and the signal is converted back to the electrical domain only at the destination. This circuit is also called “lightpath “. A lightpath is set up and taken down as dictated by the network management policy.

The main advantages of this solution are, firstly, the fact that optical bypass eliminates the requirement for expensive and energy hungry O-E-O conversions at intermediate nodes. Secondly, the all-optical routing is transparent. The transparency refers to the fact that the lightpaths can carry data at a variety of bit rates, protocols, and signal format, which enables the optical layer to support a variety of concurrent higher layers. For instance, an optical switch does not care whether it is switching a 10 Gbps Ethernet signal or a 40 Gbps OTN signal.

The arrival of new wavelength switching and routing devices, such as Optical Add/Drop Muliplexers (OADMs) and Optical Crossconnects (OXCs), has been a key enabling development of transparent optical network switching.

An OADM drops and adds a selective number of wavelengths from a WDM signal, while allowing the remaining wavelengths to pass through. Several types of OADMs exist with a range of capabilities based on the number of wavelengths they can add and drop, the ease of dropping and adding additional wavelengths, static or reconfigurable. Reconfigurability refers to the ability to select the desired wavelengths to be dropped and added on the fly. This ensures flexibility when planning the network since lightpaths are set up and taken down dynamically as needed in the network.

OADM is generally deployed to handle simple network topologies, such as linear topology or ring topologies. For large number of wavelengths or complex topologies, i.e. mesh topology or interconnection of multiple rings, OXC is deployed. As is the case of OADMs, several variants of OXCs exist, enabling to switch wavelengths, bands of wavelengths, and entire fibers.

In spite of the aforementioned advantages, all-optical approach still presents some limitations. The all-optical configuration mandates a more complex physical layer design as signals are now kept in the optical domain for a long distance. Furthermore, the number of wavelengths required in a transparent network is expected to have scalability issues since large number of wavelengths is still required within a large network to satisfy all the flows. Moreover, the rigid routing granularity could lead to severe bandwidth waste, especially when there is not enough traffic between pair nodes to fill the entire capacity of wavelengths. The

mismatch among the transmission capabilities between two nodes and the actual traffic requirements leads to a large underutilization of the resources.

The poor aggregation capability of transparent network can be overcome by combining, in a single network, transparent and opaque nodes. This allows the grooming of different traffic demands in the same lightpath and it lets some flows span over consecutive lightpaths by following multi-hop path until reaching destinations. The O-E-O conversion capability is attributed to specific nodes of the network, having a significant transit traffic load. This solution, called hybrid switching [62] [63], can improve the aggregation capability compared with the transparent solution. Thus, it is possible to achieve a more effective use of the resources which reduces not only the number of used wavelengths but also, the number of employed devices and consequently to reduce the energy consumption.

The hybrid solution could be a good trade-off between the opaque solution and the transparent one. But, ensuring an optical sub-wavelength granularity switching could further improve the energy and cost efficiency of transport network. This could be done thanks to firstly, the better traffic aggregation capabilities and, secondly, the absence of electronic traffic processing along the entire transmission path.

III.2. Sub-Wavelength switching solutions

The sub-wavelength switching solutions are proposed in order to get around the lack of flexibility of OCS solution and to benefit from the whole available bandwidth by efficiently filling wavelengths. In the ITU-T standards, sub-wavelength switching networks are referred as Sub-Lambda Photonically Switched Network (SLPSN) [4]. Huge literature concerns the time-domain sub-wavelength switching solutions presenting the accumulation of fifteen years of research. Since the description of all of these solutions seems illusory, we just present their common aspects focusing on the most promising techniques.

III.2.1. Sub-wavelength switching overview

Basically, the sub-wavelength network consists of two types of nodes: the *edge nodes* and the *intermediate nodes* (also called *core nodes*). The source side of the edge node is responsible of buffering the incoming packets and forming the optical burst by assembling

packets aiming at the same destination. The length of the obtained data burst can range from one or several packets to a short session. The burst assembly mechanism is a well-studied topic in the literature as it has a significant impact on the performance of the network. The proposed solutions can be either timer based or burst-size based depending on whether the burst is created after a given timeout or when the burst length reaches a predefined threshold. Since timer based mechanisms can result in undesirable burst lengths and burst-size based mechanisms can lead to significant latency, mixed timer and burst-size assembly mechanisms have been proposed [64] [65]. The destination side of the edge node receives optical burst, converts it into the electrical domain and then it performs the disassembly process in order to retrieve the original packets. The intermediate node is responsible only for optically switching the incoming bursts.

To perform burst transmission and switching process, most sub-wavelength switching solutions rely on a data plane and a control plane. The role of each plane strongly depends on the solution. But we can roughly summarize the tasks of each plan as follows.

The data plane has, essentially, the role of packet buffering, burst assembly/disassembly, burst transmission/reception and burst switching. While, the control plane manages the burst transmission scheduling, reserves optical resources and configures optical switches.

The control plane is the key element in all-optical sub-wavelength switching solutions since the absence of optical processing and the inflexibility of optical buffering at the intermediate node. In fact, optical buffers are generally based on Fiber Delay Line (FDL) that consists of a portion of fiber enabling to delay bursts for an only a predefined fixed duration corresponding to the burst propagation time. Hence, the control plane has the task of bringing flexibility and intelligence to the optical network. The main control functionalities can be hold in a unique or few number of control entity (a backup is needed for redundancy), it is then called *centralized*, or it can be performed locally in each node, and it is then called *distributed*. Even in a centralized control plane approach, the majority of nodes in the network hold a control unit that communicates with the centralized control entity. The role of these control units is abbreviated to receive instructions from the control entity or/and inform the control entity about some statistics concerning the node such as the state of the data queues.

The exchange of control message in the network could be done either according to an *out-of-band* approach by using a dedicated control wavelength or according to an *in-band* approach by sharing wavelengths with data. Generally, the exchange of control message is closely related to the resource reservation protocol which could proceed as one-way reservation or two-way reservation scheme regardless of the way used to transmit control messages (out-of-band or in-band). The two-way reservation is performed in two steps: a first step of requesting optical resource and a second step of confirmation or resource attribution. In this case, the data emission begins only after the reception of the confirmation. In the case of one-way reservation, the data plane does not wait for a message of confirmation and the data transmission is done after sending request message. To allocate resources, some control planes are based on an accurate mean of synchronization through the network. This synchronization is used to define slots in a cyclic process or to time-stamp the data transmission in order to avoid contention. However, other control planes are asynchronous. They don't need any synchronization between nodes.

Two main categories of sub-wavelength are considered: lossy and loss-less. Lossy solutions do not guarantee the successful transmission of data. In fact, at the intermediate nodes, bursts can compete for the same wavelength at the same time. In this case, a contention happens and, depending on the contention resolution method, one or more bursts can be discarded. In the other category, lossless solutions adopt end to end reservation of the optical resources along the path, such that contentions and, thus losses, are not possible. The reservation is usually performed by scheduling the transmission of the bursts according to well defined schemes. In the next sections, we detail the description of some lossy and lossless solutions focusing on their performance and their potential of deployment in the network.

III.2.2. Lossy sub-wavelength switching solutions

III.2.2.1. C-OBS

The *Conventional-Optical Burst Switching* (C-OBS) network architecture has been introduced in 1999 by C.M. Qiao and J.S. Yoo [3] in order to combine the best of the coarse-

grained circuit-switching and the fine-grained packet-switching paradigms while avoiding their shortcomings.

When the burst is assembled, a corresponding control packet is created and sent first on a separate wavelength to set up a connection. It is processed electronically at every core node in order to reserve the appropriate amount of bandwidth and configure the switches along the path that will be followed by the burst. According to the information carried in the control packet, each node attributes, for the arriving burst, the sufficient amount of bandwidth and the appropriate wavelength on the outgoing link.

The control packet and the burst are separated at the source as well as subsequent intermediate nodes by an offset time. At the source, the offset time is chosen larger than the total processing time of the control packet along the path. This approach eliminates the need for a data burst to be buffered at any subsequent intermediate node just to wait for the control packet to get processed.

Two different bandwidth reservation ways can be performed in C-OBS networks: Just-In-Time (JIT) and Just-Enough-Time (JET). The JIT mechanism [66] is designed to reserve resources and configure intermediate node in advance. In fact, the node configures its optical switches for the incoming burst immediately after receiving and processing the corresponding control packet. Thus, resources at the node are made available before the actual arrival time of the burst. However, in JET mechanism [67] the optical switches at a given intermediate node are configured to reserve bandwidth to the burst right before its expected arrival time and until its departure time. To do this, JET relies on the offset time and the burst length information carried in the preceding control packet.

Compared with JET, JIT is easier to implement since an accurate knowledge of the arriving time of the burst at each intermediate node is not required. At the downside, JIT leads to an underutilization of resources since wavelengths are reserved at C-OBS nodes prior to the burst arrival time. As a result, JET signaling is able to outperform JIT mechanism in terms of bandwidth utilization and burst loss probability, at the expense of increased computational complexity and a need of accurate network-wide synchronization.

The C-OBS node could receive many control packets demanding to reserve switching resources. If the reservation algorithm fails to satisfy a demand, the corresponding burst will be dropped. Among features aiming to minimize the number of dropped bursts, the C-OBS node could use Fiber Delay Line (FDL) to keep the burst in a waiting state until the availability of resources. C-OBS node could be equipped by several FDLs with different lengths, which give more choices to the reservation algorithm to manage the control packet's demand.

To ensure a reliable burst transmission, a negative acknowledgement can be sent back to the source node, which retransmits the control packet and the burst later. This retransmission mechanism can be left to the upper layer protocols such as TCP.

By processing a single control packet for a large optical payload which remains in the optical domain during its trip in the network, C-OBS is likely to bridge the gap between limited electronic processing and high optical transmission rates. But, the challenges of this technology are still related to burst loss, synchronization and control complexity.

III.2.2.2. L-OBS

In Labelled-OBS (L-OBS) [68] [69], the burst is composed of a payload section and a header section called *label*. The label carries control information required to reserve and configure optical resources for burst transmission. Bursts are asynchronous and their duration ranges from 1 µs to 100 µs with a minimum inter-burst time of 200 ns. At each node in the forwarding path, a copy of the header is extracted in order to be electronically read and processed while the burst is optically delayed by an input FDL to provide the time required for these operations.

According to the testbed described in [70] and shown in Figure 16, the extraction of the header is done by a Label Extractor (LExt) based on a semiconductor optical amplifier (SOA). The LExt is located before the input FDLs and it extracts the label clocked at a lower frequency than the payload clock and converts it in the electronic domain. Then, the electronic label is sent to the control unit which processes it and configures the switch matrix according to the result of the scheduling process.

The main challenges of the scheduling process are the avoidance of burst collision and the insurance of efficient bandwidth utilization. Consequently, the control unit has to determinate for each arriving burst the adequate wavelength that it should use and the right delay that it should wait. The scheduling process is based on Latest Available Unused Channel with Void Filling (LAUC-VF) [64] algorithm which chooses the wavelength providing the shortest delay on the burst transmission. When several wavelengths are possible, it selects the one that minimizes the gap generated between the previous reservation and the new burst reservation so as to increase the channel utilization.

According to the decisions of the scheduling algorithm, the control unit sends instructions simultaneously to the Tunable Wavelength Converters (TWCs) and the Tunable Delay Lines (TDLs) to configure them.

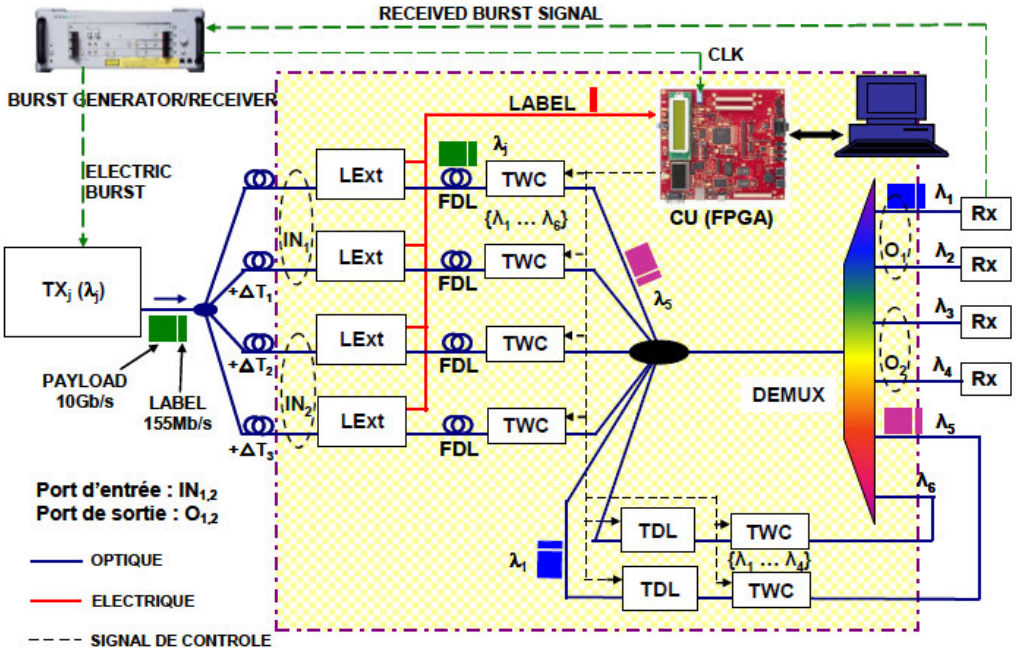


Figure 16- L-OBS test-bed

The mode of realization of L-OBS is very similar to OPS but with larger burst duration. It applies a contention resolution strategy resorting to wavelength conversion and temporal delays. The study carried out in [71] shows that this solution slightly outperforms the C-OBS in terms of burst loss probability and network resource utilization. At the downside, the same study shows that L-OBS achieves a burst loss probability of 0.1 and a capacity utilization of

0.53 for an offered load equal to 0.8. In addition to this poor performance, L-OBS uses wavelength conversion (TWC) to reduce contention. Nevertheless, all-optical TWC is still an expensive, high power consuming and immature technology.

III.2.3. Lossless sub-wavelength switching solutions

III.2.3.1. POADM

Packet Optical Add/Drop Multiplexing (POADM) [72] [73] [74] is a ring burst-switched solution proposed within the ECOFRAME project. It is partly conducted in the frame of a collaborative agreement between NTT Photonics Labs and Alcatel-Lucent Bell Labs. It is originally designed to be deployed in a WDM ring network operating at 10G with different modulation formats over 40-wavelengths. A study in [75] demonstrates the feasibility of a bit rate transparent ring on POADM. The result shows no more than 2dB penalty over the C-band for three different bit rates: 10G, 40G and 100G.

In the data plane level, POADM adopts a synchronous time-slotted approach. The slot lasts 10 µs and can transport only one burst. Each burst has a fixed duration including at least: one guard band, one preamble, one synchronization word and a payload. Dummy bursts are generated when no traffic is sent in the network to simplify the power management in the amplification stages [72]. For the control plane, POADM uses a dedicated control wavelength that can be at a different bit rate than the data wavelengths. Authors in [76] propose the structure illustrated in Figure 17 for the control message. It is composed of global control fields informing about the synchronization, the packet length, error the correcting code and the ring identifier. Furthermore, it includes at least 40 interval times representing the headers of bursts on every wavelength of the same time slot. These interval times are followed by one additional interval time for the network management and one extra interval time to transport any extra information.

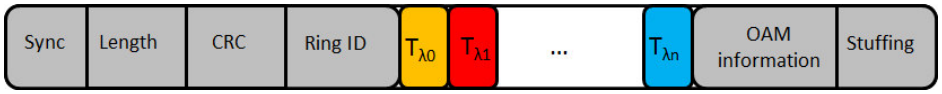


Figure 17- Control packet structure per time-slot

POADM networks are designed using two kinds of elements: Hybrid Optoelectronic Packet Router (HOPR) [77] and the POADM node. HOPR ensures the interconnection between POADM rings in the metro network architecture. When bursts transit between rings, they cross a single or several HOPRs in cascade, to bypass intermediate rings. In the HOPR, if there is no risk of contention from other incoming bursts, the burst is switched to the output port transparently (no electronic buffering). Otherwise, the burst is forwarded to an electrical shared buffer for temporary storage.

The POADM node ensures the emission, the transit and the reception of bursts. The structure of this node is illustrated in Figure 18 and is clearly described in [73]. It consists of one WDM amplifier at the input and another at the output to manage the power budget and enable the cascade of several nodes. Incoming bursts, after pre-amplification, are optically demultiplexed according to their wavelength. Besides, each burst passes through an optical coupler that splits it into two identical bursts. The first burst is dropped by a fixed wavelength receiver. Only the bursts that must be dropped are processed and the others are discarded. In the transit line, the second burst crosses an optical gate that is composed of Semiconductor Optical Amplifiers (SOA). According to the control plane instructions, SOA could be in "ON" state to let burst pass or in "OFF" state to suppress them. Afterwards, all bursts are optically multiplexed and amplified again. New bursts from the add port can by re-inserted at any wavelength using a Fast Tunable Laser (FTL). Meanwhile, the control packets are detected, to properly adopt the required switching patterns for the SOA gates and for the tunable lasers.

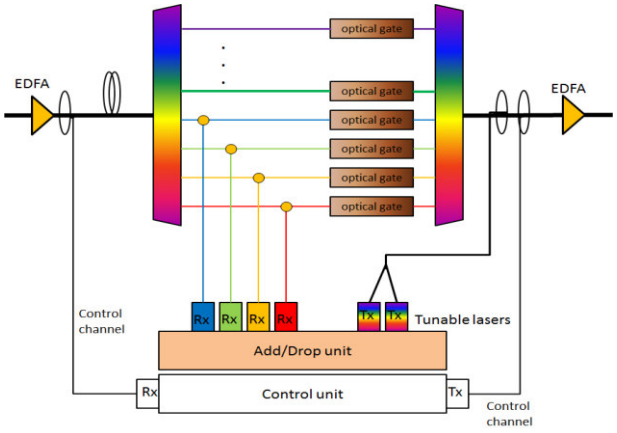


Figure 18- Packet optical Add/Drop multiplexer structure

A POADM node has a predefined number of fixed wavelength receivers and one or several number of FTL according to the network needs. The management of these devices and all the optical resources in the network is ensured by a MAC protocol that can perform centralized or distributed resource allocation.

Authors in [74] designed a centralized MAC protocol, called *Virtual circuit allocation*. In this protocol, a centralized control entity allocates resources for the entire network taking into account requests received from the edge nodes. The control entity is also in charge of interconnecting several metro optical rings to transfer the traffic between them.

A distributed MAC protocol, called SWING, is proposed in [78]. It is composed of two sub-layers: adaptation and transport. The adaptation sub-layer achieves quality of services differentiation for packets received from upper layer and it also creates optical bursts. However, the transport sub-layer is responsible for optical resources management. It combines a distributed reservation scheme, designed to ensure fairness, with an opportunistic transmission, designed to avoid wasting capacity. In the distributed scheme, if a node has more waiting bursts than the number of slots reserved to it, it seeks to reserve a slot on an available data channel by marking the corresponding control packet as it passes in the previous cycle. Moreover, the reservation done by a first node can be pre-empted by a second downstream node having a number of reserved slots lower than the first one. This preemption mechanism ensures fairness in SWING. In the opportunistic transmission, node benefits from each free slot to send a waiting optical burst to a destination that is not downstream of any node that may have previously reserved the slot.

Another distributed MAC protocol called Tag-based Enhanced Access Mechanism (TEAM) is proposed in [79]. It manages network resources using a token game mechanism. Indeed, each node holds small *token buffers*; each of them corresponds to a destination node. The generation rate of the tokens depends only on the amount of bandwidth to be reserved to the corresponding destination node. A packet is sent only if a token corresponding to its destination is available. After the emission of the packet, the token is consumed. If a token is available but data queue is empty, a *virtual packet* that does not carry any data is sent in order to maintain the reservation. If there are no tokens available while too many packets wait in the queue with high priority, packets of this class will be transmitted in the slot as Best Effort

(BE) packets, which means that they can be preempted by other packets. The preemption rules in TEAM take into account QoS differentiation. Indeed, if the length of a data queue exceeds a specific threshold, its packets can preempt BE packets. Specifically, BE packets can be dropped at intermediate nodes in order to be replaced by packets of higher priority. In this particular case, POADM loss bursts. The loss of bursts here occurs for QoS reasons and it is not related to the disability of the intermediate node to switch the arriving bursts. In order to remedy the problem of extra load due to retransmission, the dropped BE packets are stored in a *flash buffer* at the intermediate node and are retransmitted again on the first available timeslots, prior to other packets inside the same node. Note here that the value of the preemption threshold is a critical parameter that determines the efficiency of this mechanism.

Performance of POADM depends on the used MAC protocols. Authors in [80] use a totally opportunistic and distributed protocol to attribute resources. They claim that without considering bandwidth loss due to guard time between optical packets, the average wavelength occupancy of their solution is up to 80%, whereas the maximum occupancy can reach 95%. This good performance makes POADM one of the relevant sub-wavelength solutions intended to metropolitan networks.

III.2.3.2. OBTN

Optical Burst Transport Network (OBTN) [81] is an all-optical sub-wavelength-granularity transport network architecture, proposed by Huawei. It is a time-slotted solution based on an out-of-band and centralized control plane. The control wavelength carries configuration and slot reservation information from central control entity to the local control unit of each node.

In the OBTN's data plane, a wavelength is attributed to each source to transmit data to the other nodes of the network. Each wavelength is divided into equal time slots, called *Optical Burst (OB)* slots. As shown in Figure 19, OBs are grouped into frames and are time aligned with the other OBs on the other wavelengths. Two OBs occupying the same slot of time on two different wavelengths should not be destined to the same node. Accordingly, nodes are connected to each other by disjoint OB virtual paths. The OB alignment facilitates the

bandwidth provisioning and the related control. Authors in [82] suggest a frame length of 125 µs, OB length of 4µs and a guard time of 460 ns to separate between two successive bursts.

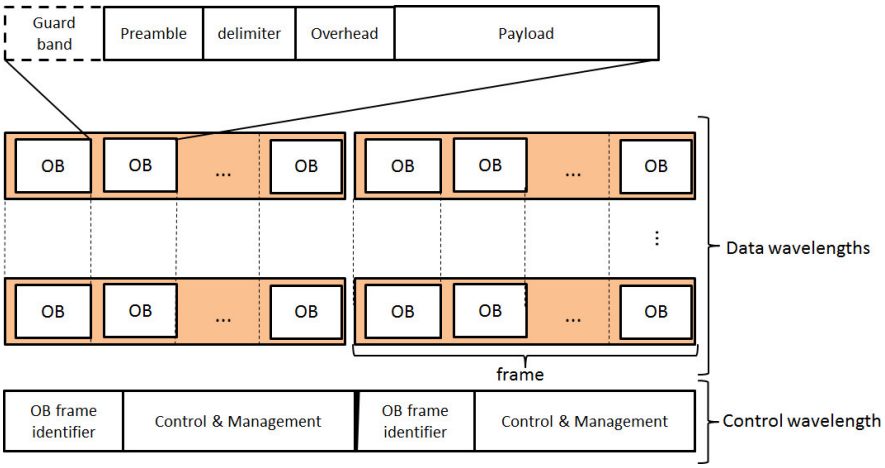


Figure 19- OB frame structure

OBTN node uses a fixed tuned laser that emits signals continuously at the source side; while, it uses a tunable burst mode receiver at the destination side. To simplify the emission and the reception process, the guard time between OBs is filled up by dummy bits. At the destination side, the arrived WDM optical burst signals are first amplified, and then split into two branches by an optical coupler. A branch continues transmission to the next node, and the other portion is fed into a Fast Optical Burst Selector (FOBS). The FOBS comprises a fast optical switch array that selects OBs destined to the node according to the information received by the control unit. The selected OBs are then fed into a BMR (Burst Mode Receiver). The structure of an OBTN node is illustrated in Figure 20.

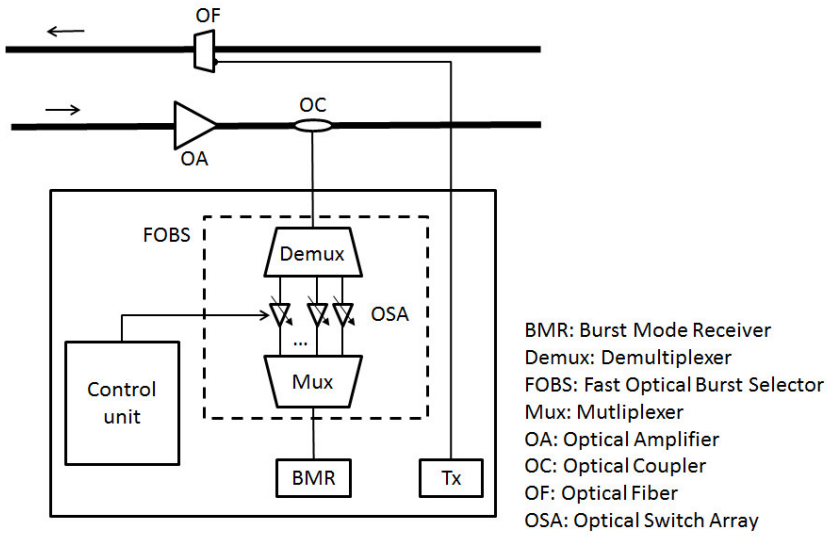


Figure 20- The OBTN node structure

OBTN can be applied in different network topologies. In the ring topology [82] [83], a Dynamic Bandwidth Allocation (DBA) scheme is employed to assign slots and configure the add/drop of the OBs automatically and efficiently. The DBA is performed by a centralized control entity that collects the bandwidth request from every node and computes the allocated OBs for each virtual connection, known as the bandwidth map. The bandwidth map is broadcasted in the network to notify nodes and configure them. In mesh topology, authors in [82] propose the same node structure as in ring topology; they only add a Fiber Delay Line (FDL) array with limited stages at each node to align OB frames coming from different input ports.

OBTN avoids the use of sophisticated or expensive components and it can easily co-exist with the present WDM networks. For instance, this solution does not need an FTL in the transmission side. However, it requires an accurate synchronization between nodes in order to ensure OBs alignment. Such a condition is difficult to provide specifically in a mesh topology and the idea of adding FDL is not recommended by operators due to the difficulties of maintenance and reparation.

III.2.3.3. OPST

The Optical Packet Switching and Transport (OPST) solution [84] [85] [86] [87] proposed by Intune Networks aims to create a robust and asynchronous packet-switching

architecture in the form of a distributed non-blocking Ethernet switch. This solution is intended to be deployed in the metropolitan area network. The data plane proceeds in two-contra-rotating optical rings that form two autonomous and redundant packet switching fabrics. The burst transmission system is designed to load balance traffic across these two optical switch fabric planes.

OPST is based on wavelength routing scheme to address packet flows. The transmitter is equipped with a Fast Tunable Laser (FTL) to rapidly switch wavelength according to the target destination, whereas, the receiver has a fixed wavelength filter, thereby the wavelength acts as the address. Hence, the OPST network is composed of a set of parallel shared media. Each of them is intended to a given destination. The access to a specific media relies on an Optical Media Access Control (OMAC) scheme inspired by the way CSMA-CA avoids collisions.

According to OMAC, incoming client packets are encapsulated as OPST frames, and then queued in a Virtual Output Queue (VOQ) on a destination and CoS basis. The scheduler composes a burst by assembling various OPST frames intended to the same destination. The most critical CoS is assembled with strict priority scheduling discipline while all other CoSs are assembled using round robin mechanism. When a burst is composed, the laser looks for a gap in the optical spectrum to transmit data. To do this, each node is equipped with an optical sensor enabling the observation of the channel state in advance phase. When the transmission system detects that the channel is free, the burst is inserted. If an upstream optical signal is detected, the burst emission is interrupted and it is resumed as soon as the channel return free. OPST node is also equipped with a FDL in order to give time to the FTL to react to a carrier-sense event and rapidly turn to a different wavelength so that burst destined to other node can be transmitted.

The data plane of OPST system can be viewed as an overlay of multiple virtual network flows that are automatically created whenever Ethernet services such as E-LINEs, E-LANs or E-TREEs are created. In order to ensure fairness amongst all active traffic and manage the distribution of resources, two dedicated control channels (clockwise and anticlockwise) are used. This control plane only needed to allocate resource by service, while the insertion of bursts is purely local without real time reservation mechanism. It works to transform the

entire network into a distributed switch. Therefore, it limits the amount of traffic on each switch plane to almost 80% of the total capacity. Furthermore, it provides two methods for forwarding traffic, namely: dimensioned resources and un-dimensioned resources. Dimensioned resources are used to guarantee bandwidth dedicated to specific services between end points, whereas the un-dimensioned resources occupy the remaining bandwidth. The control plane is comprised of three functional layers. The first layer is the scheduling layer. It describes the distribution of resources around the dual data planes. It ensures efficient use of the available channels and prevents burst collisions without the need for complicated synchronization between nodes. This feature is based on a distributed Dynamic Bandwidth Allocation (DBA) mechanism that manages the resources around the ring. The second layer is the flow control layer. It provides the functionality to create, modify and delete traffic flows in response to available resources in the ring. This function enables all nodes to discover flows of each other and its capabilities to correctly support provisioned services. The third layer is the service mapping layer which describes the mapping of network services into the traffic flow. A simplified description of some functional blocks of an OPST node is illustrated in Figure 21.

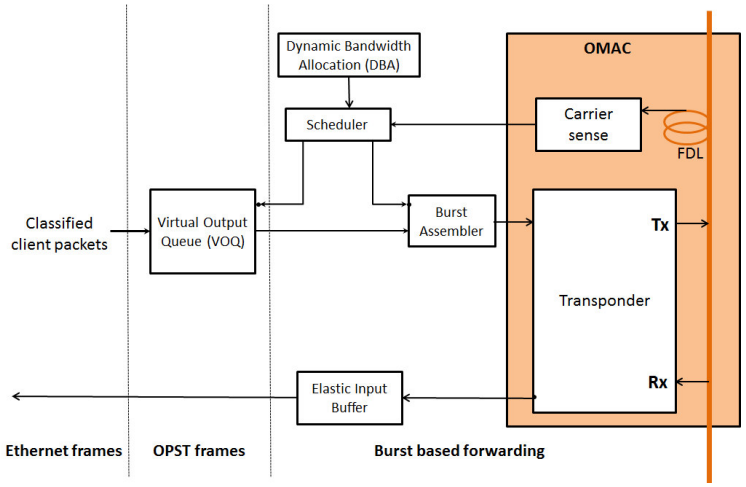


Figure 21- Functional blocks of an OPST node

The study in [84] shows that in a full mesh connectivity scenario, the capacity of the ring is between 50-60%, which means that each node is able to receive or emit up to 6 Gbps without any loss.

OPST and POADM show some similarities: both of them operate on a ring topology and they consist on a fast tunable transmitter at the source side and a fixed wavelength receiver at the destination side. However, significant differences exist and are mainly related to the way data are transmitted. In fact, burst transmission in OPST is performed asynchronously, whereas, it is synchronous in POADM. Furthermore, OPST relies on wavelength based routing unlike POADM, where one wavelength can serves multiples nodes. Wavelength based routing feature used in OPST simplifies the control plane and the burst insertion process. But, it has some drawbacks since it makes difficult to deploy multicast services and it leads to the under-utilization of available resources in the case the destination receives low load traffic. Nevertheless, despite of these drawbacks, this choice can be justified by the fact that the emission components are the most costly; so, it will be better to think to be efficient on the emitter side than on the wavelength side.

III.2.3.4. WR-OBS

Wavelength-Routed Optical Burst Switching (WR-OBS) [88] [89] combines OBS with dynamic wavelength allocation under fast circuit switching. It might be considered to be closer to dynamic circuit switching since the transmission of a burst between two edge nodes requires a dynamic set up of an end-to-end lightpath.

The lightpath establishment process is based on two-way reservation mechanism between the edge node and a centralized control entity. More precisely, client layer's packets are aggregated in the edge routers into bursts according to their CoSs and destination. At an appropriate point during the burst assembly cycle, the edge node sends wavelength request to the control node to transmit the burst. The control node sorts requests according to their CoS and schedules it using Earliest-Deadline-First (EDF) discipline so that a request which has spent more time in the queue is served earlier than the one which has spent less time there [90]. Afterwards, the control entity executes the Routing and Wavelength Assignment (RWA). Once the RWA finds an available free wavelength, the control node sends acknowledgement to the source to emit the bursts and it sends also several control messages to intermediate nodes to configure their switch. If the request exceeds the maximum delay allowed for scheduling or no free wavelength is available in the network, the request is dropped and a discard signal is sent back to the edge router. In this case, packets are not lost

but, are instead stored in the edge node buffers looking for another opportunity to be emitted. After receiving the acknowledgment, the source sends off the burst over the assigned lightpath through the core network without intermediate optical processing. Once the whole burst has been transmitted, the lightpath is released and becomes available for subsequent connections. Figure 22 shows the request server architecture in the control entity.

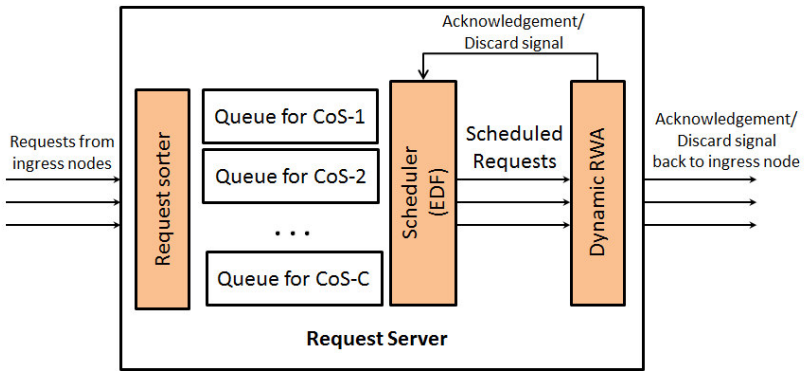


Figure 22- Request server architecture

The established lightpath in WR-OBS is held only for the burst transmission time plus end-to-end propagation delay. If insufficient wavelength holding time is reserved to the source due to erroneous prediction of the burst size, the burst can be sent only in part and the remainder of the burst is dropped. Therefore, the burst size prediction is an important mechanism in WR-OBS solution that has a significant impact on the resultant network performance. The burst size prediction is done by the control entity based on packet buffer filling statistics. Indeed, the request message sent by the source to control entity contains information about the amount of data that have been already accumulated. Since the burst assembly process at the source continues until the reception of the acknowledgment, the control node has to estimate the amount of traffic received by the source in the time interval between the emission of the request and the reception of acknowledgement. For this purpose the control entity uses a feedback control loop based on the statistics that it collects during the previous connections [91].

According to the study done in [92], authors demonstrate that a traffic load of up to nearly 70% of the total link capacity can be carried by the WR-OBS network while satisfying the QoS requirements. Despite of this acceptable performance, this solution suffers from an important control overhead since several control packets have to be sent before each burst

transmission to request and confirm the lightpath set-up and also to configure the switches along the lightpath.

III.2.3.5. TWIN

Time-domain Wavelength Interleaved Networking (TWIN) was originally invented by Bell Labs [5] [93]. It is a cost-effective network architecture that can provide flexible connectivity using passive optics in internal nodes. Indeed, a particular wavelength is attributed to each edge node to receive its data. When a source has a burst to send to a given destination, the source tunes its laser to the wavelength uniquely assigned to that destination for the duration of the burst. The intermediate nodes steer optical signals passively from their inputs to their outputs based on the color of the burst. Thus, the virtual topology of TWIN can be viewed as overlaid optical multipoint-to-point trees. Each of these trees has a unique color and it is associated to a unique destination. To perform automatic discovery of resources, routing and signaling, TWIN adopts a separate control plane by allocating a dedicated wavelength for this purpose. Figure 23 shows an example of TWIN architecture.

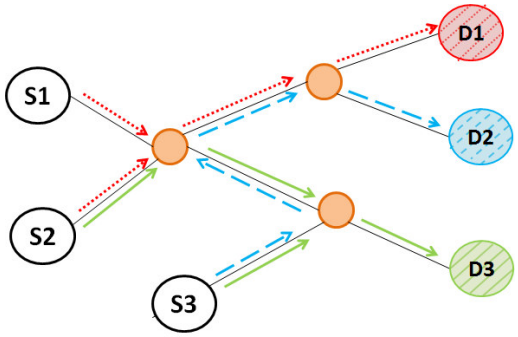


Figure 23- TWIN concept

According to this architecture, the complex processing functions are pushed to the network edge such that the network core only has to deal with an optical forwarding layer. Edge nodes utilize burst-mode receivers and fast tunable lasers to emulate fast switching of data in the core. Whereas, intermediate nodes consist of a passive wavelength switches, i.e. Wavelength Selective Switch (WSS), capable of merging and routing incoming wavelengths to the appropriate outgoing ports. The cross-connect configuration stays at very long time scales since reconfiguration is only needed when a failure occurs or a new connection requires a new branch of a tree to be created.

The fact that all sources share the same medium to reach a specific destination leads to possible collisions at each merging point of the tree. To resolve this problem, TWIN relies on a complex scheduler to coordinate sources transmission. To support both *synchronous* and *asynchronous* traffic, TWIN adopts both a *centralized* scheduler [94] and *distributed* scheduler [95] [96] respectively. The transmission to a given destination is organized in repetitive cycles. A cycle consists of a predefined number of slots and each slot carries exactly one burst. The purpose of the scheduler is to assign, in each cycle, the appropriate slot(s) to source-destination pairs to avoid collisions. Each cycle is divided into two periods. Each period is managed differently by one of the scheduler (centralized or distributed). Boundary of periods is flexible and negotiated between schedulers.

The centralized scheduler is performed in a particular control point within the network. The control point gathers all necessary information (e.g., traffic demand matrix) and processes it in a relatively long time interval. Then, it computes the slot allocations to each source-destination pair effectively and it sends it to the edge nodes via an out-of-band control channel. Authors in [94] propose a-generic approach based algorithm called TWIN Iterative Independent Set (TIIS) to perform a centralized scheduler. The algorithm is a heuristic approach to compute the minimum number of slots needed to complete the transmission of the entire demand matrix taking into consideration the maximum difference in propagation times. The algorithm, then, executes many iterations in order to find the best assignment for burst timeslots.

The distributed scheduler is suitable for asynchronous traffic with dynamic bandwidth requirements. For faster response time, a control point is located in each destination and it independently attributes slots to sources that are transmitting to it, basing only on source resource requests. The main drawback of this scheduler is the fact that a source may receive multiple grants that call for it to transmit simultaneously, which can create conflict. The algorithm proposed in [96] takes the form of a congestion control protocol where slot assignments depend on feedback received regarding previous collisions.

In [97], authors introduce a novel variant of TWIN that retains its main characteristics. This variant is based on assigning a wavelength per source (instead of the destination) in such manner that source becomes the root of a multipoint-to-point lightpath shared by all

destination nodes. Then, the tunable transmitters and the fixed-wavelength receivers are replaced, in the new variant, by fixed-wavelength transmitters and tunable receivers. This enables to take advantage of coherent detection which enables higher rate optical reception with fast switching between wavelengths. Similar scheduling algorithms as those used in the original version of TWIN could be applied to this variant taking into account, of course, the reverse structure of the trees.

TWIN concept looks interesting in terms of the fast switching and the avoidance of optical buffers in the intermediate nodes. It also enables self-routing in the network core as packet-forwarding relies on the wavelength rather than label/address lookup. Nevertheless, TWIN suffers from the complexity of scheduling algorithms. Moreover, the assignment of a determinate set of wavelengths to each egress node may lead to scalability issues and to fiber link underutilization due to the lack of wavelength reuse.

Since the original TWIN did not scale well since the number of nodes was limited by the number of available wavelengths, authors in [98] [99] propose the so-called *TWIN with Wavelength Reuse* (TWIN-WR) to circumvent this constraint. Unlike in TWIN, a source node in TWIN-WR may not be able to send traffic directly to any destination node in an optical single hop, resulting in multihopping via intermediate electrical gateways. TWIN-WR firstly assigns W wavelengths to the N nodes with the objective of maximizing the throughput of direct traffic. Besides, it creates the virtual topology by designing the multipoint-to-point tree for each destination, taking into account physical layer constraints. In the virtual topology, a cycle of lightpath is set up between nodes having the same wavelength. The same centralized and distributed scheduling algorithm used in TWIN could be used in TWIN-WR to coordinate the tunable lasers. The main advantage of wavelength reuse is the reduction of the number of required wavelengths to cover all the network demand. But, this is done at the expense of increasing the number of hops required to cross the network and also the dependencies between receivers.

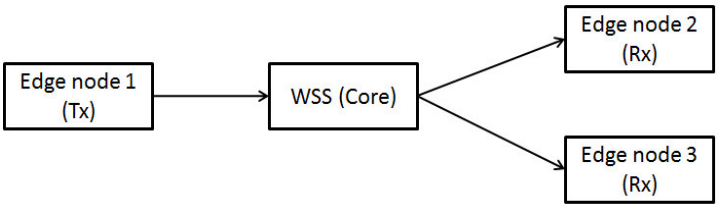


Figure 24- First TWIN test-bed prototype

A first test-bed of TWIN is realized in Shanghai Jiao Tang University in 2007; it is described in [100]. This test-bed, as shown in Figure 24, does not contain a control plane. It includes only one source edge node, one intermediate node employing a 1x2 WSS, and two destination edge nodes, each consisting of a photodetector. At the source side, an FPGA generates parallel bursts to be transmitted at 125 Mbps. Bursts intended for different destinations are put in different queues. The parallel data is then sent to a serializer/deserializer that outputs 1.25 Gbps serial data. Besides, bursts are forwarded to the transmission unit that is composed of two tunable lasers. The first laser emits data bursts on the adequate wavelength, while, the second laser emits dummy bursts that fill empty slots on the other wavelength. Dummy bursts ease the clock recovery and relax the requirements on the reception. At the output of the laser, each data burst lasts 1.55 µs (1948 bits) with 80 ns guard time. The fact that the dummy bursts are integrated in the emission side and are kept in the network along their way to the destination causes a full occupation of the wavelength. This conception makes impossible to upgrade the system by integrating a control entity and a second source emitting to one of the existing destinations.

III.3. Discussion

In this chapter we have presented several optical switching technologies, which we classified into two main categories: wavelength switching and sub-wavelength switching solutions. The wavelength based switching solutions are currently used by telecommunication operators. They evolved from opaque to transparent switching. The transparent switching eliminates O/E/O conversion in the intermediate nodes at the expense of absence of aggregation in these nodes. The coarse granularity of attributed resources, equal to the transmitter's capacity, generates the underutilization of available bandwidth. Then, switching sub-wavelength entities inside the channels seems interesting to benefit from the whole

available bandwidth. Among the possible solutions, the time-domain sub-wavelength switching represents a good option as it performs flows aggregation without resorting to electronic and its O/E/O (Optical/Electrical/Optical) conversion interfaces.

In this chapter, we have carried out a deep study of the SLPSN solutions. Hereby, we focus on the main characteristics of these solutions and we highlight their important common features. We classify the different SLPSN solutions according to a main criterion which is the presence or the absence of possible data loss during the trip of the burst through the network. Hence, we distinguish lossy solutions from lossless solutions.

In lossy solutions, the intermediate node locally performs the avoidance of burst collision. In some cases, the decision taken by these nodes consists in dropping a burst, which consequently generates the loss of all the packets that compose it. This trouble affects the network performance and throughput accuracy since it is unpredictable and it leads to high packet loss ratio that could exceed 10-4 in some cases. Compensation methods, such as contention resolution, retransmission, wavelength conversion or correcting codes, often cause degradation of delay jitter and throughput. For instance, the retransmission of bursts in a network where distance between nodes is superior to 100 km generates a significant delay. Besides, compensation methods often improve packet loss ratio at the expense of having more complexity, specifically at the intermediate node. Given the quality of service requirements associated to the transport network, lossy solutions seem inoperative and it is necessary to move towards lossless solutions.

In lossless solutions, the benefits of transparent grooming are fully obtained since all-optical switching is performed without any burst collision. Thanks to a robust control plane the transmission of burst is managed such that contentions at intermediate nodes are avoided. Throughout the literature, many approaches are possible to design the control plane. Some solutions are based on centralized allocation of resources, while others use a distributed approach where many nodes should coordinate to control the network. Furthermore, the exchange of control message could be in-band using the same wavelength as data or out-of-band by using a dedicated control wavelength. Moreover, the reservation protocol may proceed as one-way reservation or two-way reservation scheme. The choice of control plane mechanism could be driven by the network topology. In a mesh topology, the control plane

has to accurately know the propagation time between the different nodes in order to perform its resource allocation algorithm. Moreover, the solution requires a perfect synchronization between nodes. However, in a ring topology, the implementation can be synchronous or asynchronous as it is the case of OPST of Intune. In Table 1, we summarize the characteristics of the solutions studied in this chapter according to the aforementioned criteria.

	Burst loss (Yes, No)	Mesh/Ring (M,R)	Synchronous/ Asynchronous (S,A)	Distributed/ Centralized (D,C)	In-band / Out-of-band (I,O)	No/One way/ Two way reservation (0,1,2)
C-OBS	Y	M	A	D	I/O	1
L-OBS	Y	M	A	D	I	1
POADM	N	R	S	C	O	0,2
OBTN	N	M/R	S	C	O	2
OPST	N	R	A	D	O	2
WR-OBS	N	M	S	C	O	2
TWIN	N	M	S	D/C	O	2

Table 1- Classification of different SLPSN solutions

Lossless solutions seem more adapted to operational networks than lossy ones. At the matter of fact, SLPSN technologies are currently evolving into this trend under the influence of operators and manufacturers. For instance, lossless trend characterizes POADM of ALU, OBST of Huawei and OPST of Intune. However, it is difficult to take a firm decision on the best lossless solution. The evaluation criteria should take into account the use case where the solution will be used (topology, size and type of network, traffic matrix etc…). The more the solution is flexible, the more it can cover use cases.

In our study, we focus on TWIN paradigm since it is a lossless solution and it is designed to be deployed on a mesh topology. It seems interesting because of its node structure simplicity and its bandwidth flexibility. From a node structure point of view, only edge nodes perform electronic buffering, while the intermediate nodes consist of passive optical components and operate at full optical capacity without any electronic processing. Compared with conventional OBS, optical buffering and fast optical switching at each node are not also needed. In terms of bandwidth, TWIN supports unpredictable traffic patterns and manages potential traffic variation by changing the amount of bandwidth allocated to a given source/destination traffic without actually changing any physical connection.

Chapitre IV. **TWIN Medium Access Control**

Since more than fifteen years, many optical burst technologies have been presented either theoretically or experimentally. One of the main challenges of OBS solutions is to avoid burst collision at each node of the network. Unlike electronic packet processing, where buffering is used to avoid conflicts, optical burst networking requires bufferless operation at intermediate nodes, because photonic memories don't still exist as a mass-produced component. One of the main drawbacks of classical OBS solutions, such as C-OBS or L-OBS, is the collision of bursts going to the same destination at the same moment which leads to the loss of losing some of them(contention). In the state of the art concerning OBS technologies, some lossless solutions have been proposed based on the idea of providing a simple and passive switching at the intermediate nodes. Especially, the *Time-domain Wavelength Interleaved Networking* (TWIN) solution developed in Bell Labs is one of the promising sub-wavelength solutions. As seen in the previous chapter, the main idea of TWIN is the attribution of one (or more) wavelength(s) to each destination node of the network to receive traffic from the other nodes. Burst collisions are avoided by a control plane such that burst, emitted by an edge node at a specific moment, optically bypasses all the intermediate nodes and reaches its destination without being buffered or receiving any electronic processing along its path. These features could fulfil the high performance required in carrier networks. In addition to that, it could provide low energy consumption thanks to the all-optical switching.

Compared with traditional OBS solutions, TWIN does not require wavelength conversion or the use of a header. Furthermore, it does not need fast optical switching mechanism in the

intermediate node as it is the case of some sub-wavelength switching solutions like POADM. It is applicable to both mesh and ring topologies and relies on a juxtaposition of multi-point to point tree networks. Each tree seems to be similar to a PON tree used in access networks. The destination, at the root of the tree, is like the OLT and the source nodes, at the leaves of the tree, are like ONUs. However, differences between these two technologies exist. They are mainly related to the fact that each source node in TWIN sends data to many destinations. This means that the multi-point to point trees of TWIN are not independent relative to each other. This characteristic leads to additional constraints concerning resource attribution and hence, to the need of more efficient and complex control plane to manage the sending of bursts between each source-destination pair.

In this context, two main control schemes might be defined as already described in the previous chapter: the centralized and the distributed schemes. In the centralized scheme, the resource allocation is done from a centralized control entity (CE) that has access to the complete network state, including network topology and requests from sources. In the distributed scheme, the control is shared between several nodes.

In this chapter, we propose new solutions for the control/management plane of TWIN technology, based on four main mechanisms: the signaling, the traffic estimation, the resource allocation and the slot assignment. As a first study, we compare, by simulation, the distributed control plane and the centralized one in terms of bursts end-to-end delay, jitter, queue length and total bandwidth utilization. Afterwards, we focus on the centralized control plane. More particularly, we emphasize the resource allocation mechanism. Therefore, we propose four different algorithms for this mechanism and we compare their performance. The target of the comparison study is to define the best centralized algorithm. The comparison is performed in a metropolitan scenario where the distances between nodes are taken in the range of few hundreds of kilometres, which is typically one order of magnitude larger than PONs. We also consider implementation constraints such as the impact of wavelength switching time of the tunable lasers on the guard time between two successive bursts, the synchronisation uncertainties and the clock and phase recovery at receiver side.

IV.1. Motivation

One of the main advantages of TWIN is its ability to perform an all-optical burst switching at the intermediate nodes using passive components. Thus, all the power consuming devices are pushed to the edge of the network. In order to assess this feature of TWIN and understand in which conditions it is interesting, we perform a dimensioning study [6] in which we retrieve the number of transponders required to sustain a metropolitan-like use case scenario. In this study, we compare three sub-wavelength switching technologies (C-OBS, POADM, and TWIN) with legacy circuit switching technologies (opaque, transparent and hybrid). Transponders are a key element in the design of transport optical solutions. The number of required transponders has an impact on the capital expenditure (CAPEX) and the power consumption of the technology. Hence, this type of study can evaluate the energy efficiency of TWIN compared with other optical transport technologies.

We present the dimensioning results in a 10-node fully connected in a unidirectional ring network. The traffic is uniformly distributed among the nodes of the ring and the capacity C of Tx and Rx is set to 10 Gbps. We aim to retrieve the total number of Tx and Rx per node in order to achieve a desired load per flow. We define load per flow as the total amount of traffic successfully transmitted between two edge nodes.

The dimensioning for the opaque and for the transparent switching technologies is retrieved analytically. In the opaque case, when an optical wavelength passes through a node, it is received by the Rx, converted into electrical domain in order to add or drop data and then it is retransmitted again by the Tx. The number of flows F, transiting on each link of a unidirectional N-node ring network, is given by the Equation IV-1.

$$F = \frac{N(N-1)}{2} \qquad \textbf{Equation IV-1}$$

We define ℓ as the traffic arrival rate on Gbps of each flow and C as the capacity on Gbps of Tx (or Rx). Then, the number of Tx (or Rx) in each node (n_{opaque}) is given by the Equation IV-2.

$$n_{opaque} = \left\lceil F.\frac{\ell}{C} \right\rceil \qquad \textbf{Equation IV-2}$$

⌈x⌉ means the nearest integer larger than x

In the transparent case, each node is sending directly the traffic to all the destinations. The number of needed Tx (or Rx) depends only on the amount of emitted (or received) data. So, the number of Tx (or Rx) required per node ($n_{transparent}$) is calculated as follows:

$$n_{transparent} = \left\lceil \frac{\ell}{C} \right\rceil . (N-1) \qquad \textbf{Equation IV-3}$$

In the case of hybrid circuit switching, it is necessary to perform the design of the network in order to choose which lightpaths to be established and how traffic demands are routed into the lightpaths. We utilize for that a meta-heuristic based on a genetic algorithm proposed in [101] in order to minimize the total number of Tx and Rx in the network.

The number of Tx and Rx, for the sub-wavelength switching technologies, is determined using simulations. We use the discrete event simulator *OMNeT++* [102] as the network simulation framework. We consider an opportunistic MAC layer for POADM and a distributed control plane for TWIN, where each destination performs resource reservation for its related source nodes. Since there are no losses in TWIN and POADM, we dimension the network considering, at each source, a traffic arrival rate per destination ℓ equal to the desired load per flow. In the case of C-OBS, we choose to dimension the network by increasing the arrival traffic rate until the lost bursts are compensated and the desired load per flow is reached. In this way, we emulate the retransmission and we take into account the additional resources required for it.

For simulation, we consider fixed-size bursts of b=*5Kbytes*. At each node, the bursts intended for a given destination arrive according to a Poisson process with arrival rate equal to ℓ/b. The time slots have a fixed duration equal to the duration of a burst plus a guard time equal to 5%, in order to take into account laser tuning time and synchronization accuracy issues.

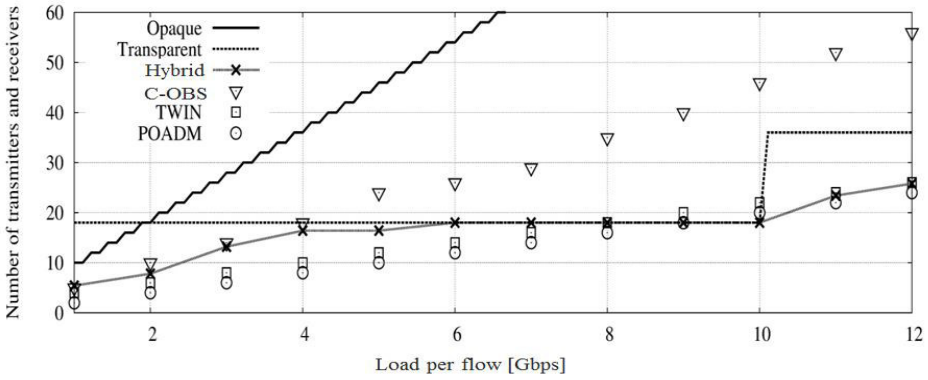


Figure 25- Number of transmitters and receivers per node vs flow per node

The total number of Tx and Rx required by a node to achieve a given load per flow is depicted in Figure 25. The figure shows that opaque circuit switching requires limited resources just for very low traffic loads, while, as soon as the traffic load grows, the required resources steeply increases. Instead, transparent switching achieves interesting results just when the flow per node is close to the Tx capacity. Indeed, the drawback of transparent switching is that it needs a significant initial number of Tx and Rx, due to its poor aggregation capacity.

As expected the hybrid switching solution is always performing better with respect to the opaque and transparent cases. Indeed, it is able to choose, depending on the traffic load, which is the best trade-off between opaque switching and direct optical transmission. The lossy OBS performs well only at low loads. Indeed at high loads, the over-dimensioning, required to recover from the burst losses, has a real detrimental effect on the dimensioning results. Thus, the absence of coordination for the transmission seems to have no particular advantages, apart that no synchronization is required.

On the other side, TWIN and POADM are performing better than the hybrid circuit switching at low loads, despite of the guard time between bursts and the distributed nature of their scheduling. However, in the high load case, TWIN and POADM perform very close to the hybrid solution. This small degradation of performance of sub-wavlength solutions is mainly related to the waste of transmission capacity due to the guard time between bursts and the performance of the control plane. Using other sophisticated control planes could improve the performance but curves keep the same general shape.

From this preliminary study, we have shown that in a low and uniformly distributed traffic scenario, optical sub-wavelength switching technologies can reduce the number of required Tx and Rx in the network with respect to legacy circuit switching. Although we have used primitive control planes for the lossless sub-wavelength solutions, their performance is interesting. By considering more sophisticated control planes, these results could be considerably improved.

TWIN is topologically more flexible than POADM since it is intended to be deployed in a mesh topology. Therefore, we will focus on this technology in the rest of this study. Specifically, we will propose new control planes and we will compare them.

IV.2. TWIN control plane overview

In TWIN network, the destination side of each node is assigned to a multipoint-to-point tree for reception. The reception trees are pre-provisioned at distinct wavelengths and overlaid on the physical network as shown in Figure 26. However, the source side of the node is related to all the trees. Therefore, each source is equipped with a fast-tunable laser. When a burst is ready to be sent to a given destination, the source tunes its laser to the wavelength uniquely assigned to the corresponding tree for the duration of the burst.

Each intermediate node performs self-routing of optical bursts to the adequate output port based solely on the wavelength of the burst. No label/address lookup processing is needed in forwarding bursts from one node to another, thereby making the network core transparent and simple. Intermediate nodes are pre-configured so that any incoming optical signal of a given wavelength will be routed to the appropriate output of the node.

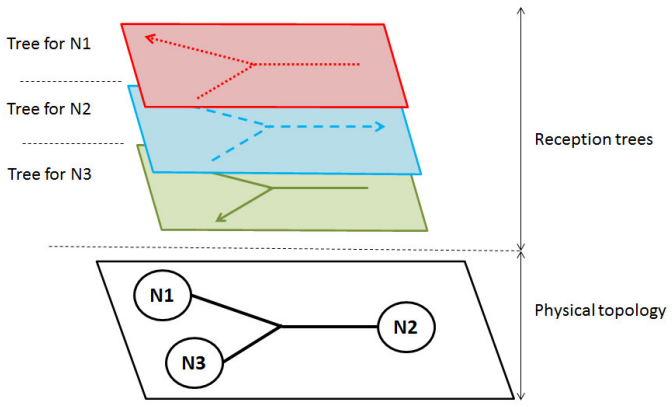


Figure 26- Overlaid trees for burst transfer in TWIN

The main task of the control plane in TWIN architecture is to efficiently manage source emissions while burst collisions are avoided both in the core and in the destination nodes.

IV.2.1. Control plane mechanisms

For both centralized and distributed control planes, we distinguish four different mechanisms to design the control/management plane: the signaling, the traffic estimation, the resource allocation and the slot assignment. As depicted in Figure 27, the signaling mechanism performs the exchange of control messages between the CE and the control units (CU) at the source side of the node. The traffic estimation and slot assignment mechanisms are implemented in the source side, whereas resource allocation is implemented in the CE side. In the distributed scheme, control entities are located in each destination node, while, in the centralized schemes, the CE is a unique and a particular node of the network. These mechanisms composing the control plane of TWIN will be further described in the section IV.3 of this chapter.

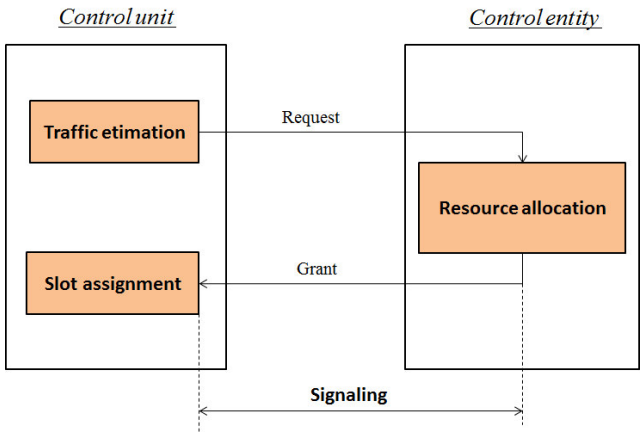


Figure 27- Control plane mechanisms

Each of those mechanisms is executed at a specific time during the control process. Therefore, a perfect synchronization is needed to ensure the reliability of the system. The slightest timing mistake could lead to a huge number of burst collisions and then the loss of enormous amount of data.

IV.2.2. Time repartition model

In our proposition in [7], control plane is organized by repetitive cycles that we call *control cycle*. The duration of a control cycle is common to all destinations and it exceeds the

duration of the round-trip time of the most distant couple of CE-CU pair in the network. As shown in Figure 28, a control cycle consists in a predetermined number *c* of *data cycles*. Each data cycle is divided into a predetermined number *n* of slots. The time slot can carry only one single burst and adjacent bursts are inter-spaced by a guard time in order to take into account implementation factors such as time-of-day synchronization errors and component switching times. All the data cycles of a given control cycle use the same allocation configuration. This configuration changes from a control cycle to another. This feature enables to have a flexible control plane that can react according to different time scales depending on the duration of the control cycle. Moreover, this feature separates the time scale of the control plane (control cycle duration) from that of the data plane (data cycle duration).

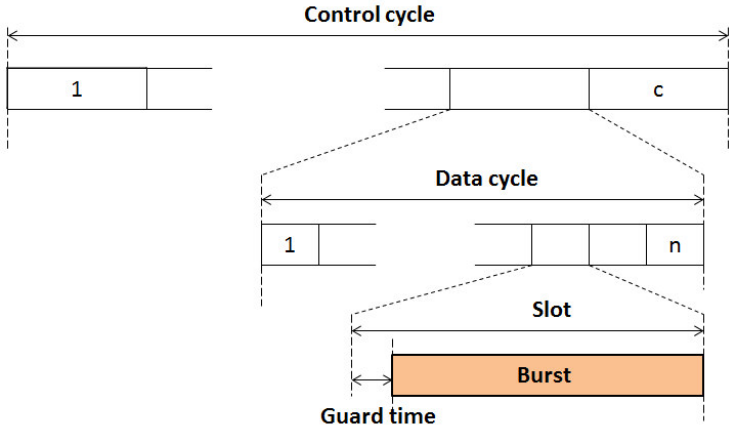


Figure 28- Time repartition of the control cycle

In the source i, the start time T_{ij} of the current control cycle of a given destination j is calculated as follows:

$$T_{ij} = T - \delta_{ij} \qquad \textbf{Equation IV-4}$$

Where T represents the start time of the current control cycle at the destination side. Here, we assume that control cycles start at the same time T in each destination. δ_{ij} is the propagation delay between source i and destination j. Then, the start time of the p-th slot of the m-th data cycle ($t_{i,j}^{m,p}$) is calculated as follows:

$$t_{i,j}^{m,p} = T + (m-1).\Delta_d + (p-1).\Delta_s - \delta_{ij} \qquad \textbf{Equation IV-5}$$

Where Δ_s is the duration of one slot and Δ_d is the duration of one data cycle.

Let's consider that a source i emits traffic to two different destinations j and j', and $\Delta = \left| t_{i,j}^{m,p} - t_{i,j'}^{m',p'} \right|$ where, $t_{i,j}^{m,p}$ and $t_{i,j'}^{m',p'}$ are the start times of the slots p and p' dedicated respectively to destination j and j' as viewed in the source i side. If Δ is a multiple of Δ_s we say that destinations j and j' are *slot-aligned* in the source i otherwise j and j' are considered *non-slot-aligned* in the source i. Figure 29 illustrates these two concepts. If $0 \leq \Delta < \Delta_s$ and source i is equipped with only one transmitter, the two slots p and p' cannot be used at the same time. In this case, we say that slot p and slot p' are *overlapped* in the source i. In the slot-aligned case, a given slot is overlapped with only one other slot per destination. However, in the non-slot-aligned case, a given slot is overlapped with two slots per destination.

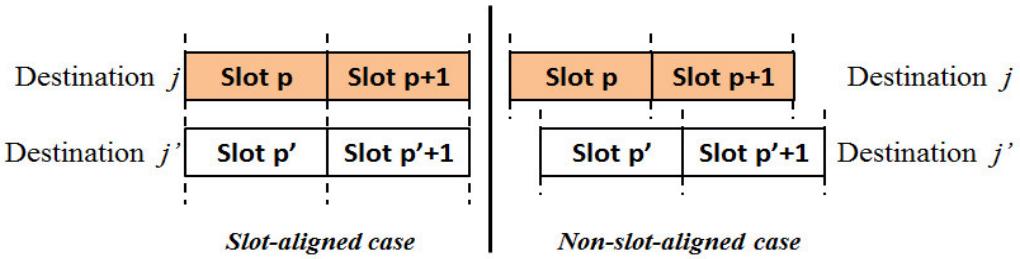


Figure 29- Slot-alignment vs non-slot-alignment in the source side

IV.3. Description of the control plane mechanisms

As mentioned in section IV.2.1 of this chapter, we divide the control plane into four main mechanisms interacting between them.

IV.3.1. Signaling mechanism

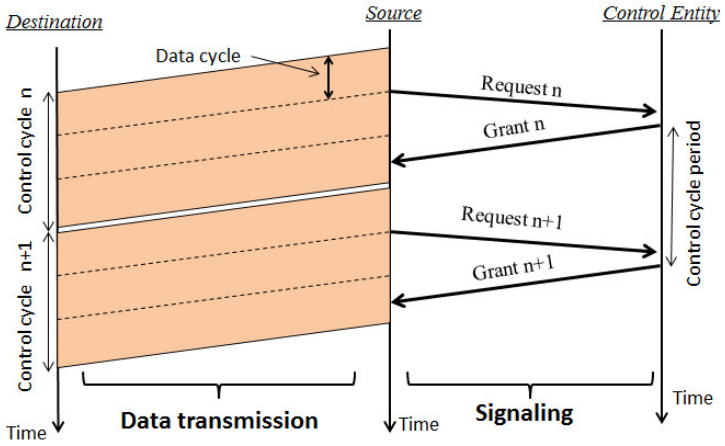


Figure 30- Signaling mechanism

The signaling mechanism, illustrated in Figure 30, has the task to ensure the exchange of control messages between the CE and the CU at the edge node. The CU at the source node makes an estimation of its traffic and sends the number of required slots to the CE via a *request message*. By taking into account requests coming from source nodes, the CE calculates the resources that can be allocated to each source without generating any burst collision in the network. The CE attributes slots to sources and sends them the indexes of those slots within the data cycle, via a *grant message*. Thus, the grant provides a bursts emission pattern for the source that it follows during all the data cycles of the next control cycle. Of course, grant messages should arrive to the source before the start time of the next control cycle. Otherwise, the source continues to use the obsolete bursts emission pattern, which could lead to collision with bursts emitted by other sources and using new patterns.

IV.3.2. Traffic estimation mechanism

In this process, sources estimate the number of required slots during each data cycle of the next control cycle. They communicate this traffic estimation with the CE via the request message. The purpose of the request is to ensure up-to-date information at the CE so that, firstly, the service of the coming packets and, secondly, the evacuation of waiting packets in the queues for each destination are properly achieved. The determination of the amount of resources to request for the next control cycle is based on statistics collected during the previous control cycle. We calculate it as a function of the queue size and the received packets in the previous control cycle.

To explain the traffic estimation mechanism that we propose, we consider firstly a burst level system where we assume that arriving and the queued data are in the form of bursts. In this case the traffic estimation is done as follows:

At the end of each data cycle k, source i counts the number of burst intended to destination j that arrived during that data cycle ($\lambda_{ij,k}$) and it also takes the size of the queue (number of bursts in the queue) dedicated to j ($q_{ij,k}$). Then, it computes the mean arrival rate of packets intended to j during a data cycle ($\overline{\lambda_{ij}} = \frac{\sum_{k=1}^{c} \lambda_{ij,k}}{c}$) and the mean length of queue related to j ($Q_{ij} = \frac{\sum_{k=1}^{c} q_{ij,k}}{c}$) where c is the number of data cycle per control cycle.

In order to reduce queue length fluctuation and guarantee the stability of the system, the emptying of queues should be done progressively (during several data cycles). For this reason, we introduce a *damping factor K* when estimating the amount of resources to request for the next control cycle. So, the number of bursts remained in the queue since the previous control cycle and that have to be served in each data cycle of the next control cycle are equal to:

$$\overline{Q_{ij}} = \frac{Q_{ij}}{K} \qquad \textbf{Equation IV-6}$$

The number of slots r_{ij} required to serve bursts from source i to destination j is finally computed by the following equation:

$$r_{ij} = \left\lceil \overline{\lambda_{ij}} + \overline{Q_{ij}} \right\rceil \qquad \textbf{Equation IV-7}$$

$\lceil x \rceil$ means the nearest integer larger than x

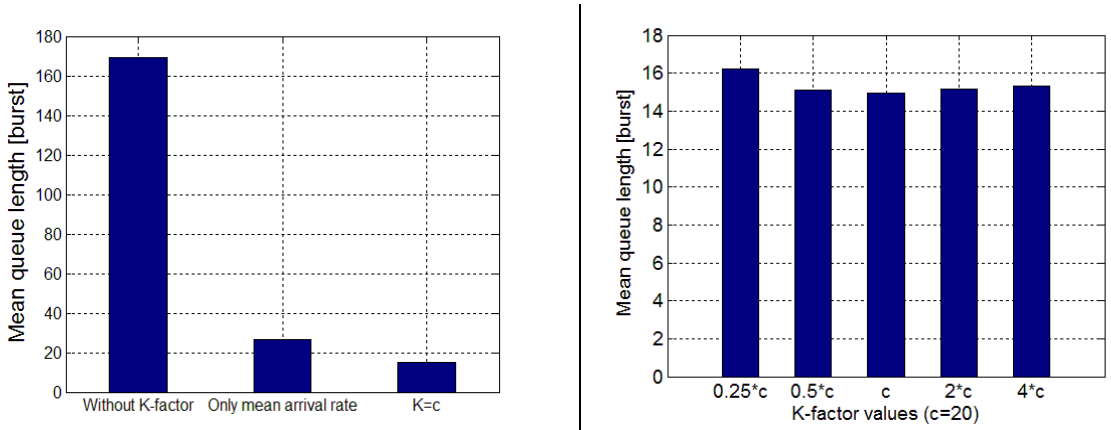


Figure 31- Determination of the damping factor

We carry out a simulation study to estimate the value of the damping factor. Results of this study are depicted in Figure 31. In (a), we compare three estimation methods. In the first one we do not consider the damping factor, in other words, we take *K=1* in the Equation IV-6. In the second method, we consider only the mean arrival rate of bursts ($r_{ij} = \left\lceil \overline{\lambda_{ij}} \right\rceil$) which means that *K* is equal to infinity. In the third method, *K* is equal to *c* (the number of data cycles per control cycle). By considering K equal to 1, the mean length of queue is significantly important comparing with the other methods. The third method enables the

decrease of the mean length of queue by almost 45% respect to the second method that is based only on the mean arrival rate of bursts. In (b), we compare different value of K. Result shows that taking K equal to the number of data cycles per control cycle ensures the less queue length.

Now, we consider a packet level system where data arrives to the system and are queued in the form of packets. The packet level system is similar to the burst level one but just we have to take into account the burst assembly process. In this case, the source *i* counts, at each data cycle *k*, the total size of packets arriving to the system and intended to destination *j*. we refer to this quantity as ($\lambda'_{ij,k}$). Meanwhile, the source *i* takes at the end of each data cycle the total size of packets in the queue and dedicated to *j* ($q'_{ij,k}$). Then, the mean total size of packet arriving to the system during a data cycle and intended to *j* is: $\overline{\lambda'_{ij}} = \frac{\sum_{k=1}^{c} \lambda'_{ij,k}}{c}$ and the mean total size of packets in the queue and related to *j* is equal to $Q'_{ij} = \frac{\sum_{k=1}^{c} q'_{ij,k}}{c}$. By introducing the damping factor *K*, we get:

$$\overline{Q'_{ij}} = \frac{Q'_{ij}}{K} \qquad \textbf{Equation IV-8}$$

Then, by taking into account the maximum burst size per slot *b*, the number of slots r_{ij} required to serve bursts from source *i* to destination *j* is equal to:

$$r_{ij} = \left\lceil \frac{\overline{\lambda'_{ij}} + \overline{Q'_{ij}}}{b} \right\rceil \qquad \textbf{Equation IV-9}$$

IV.3.3. Resource allocation mechanism

The resource allocation consists in reserving slots of time for a given source to send its burst for a given destination. Because of the tree structure, bursts that are timed not to collide at the destination cannot collide anywhere else in the network. This characteristic of TWIN alleviates the complexity of the resource allocation functionality. As a result, the main concern in the distributed scheme is to avoid the burst collisions in the destination receiver. In the centralized scheme, the CE deals with one more issue: the avoidance of burst collision in the destination nodes and the avoidance of *slot blocking* in the source nodes. A resource blocking happens when source receives multiple grants asking it to transmit during

overlapped slots towards multiple destinations. If the number of transmitters of the source is smaller than the number of overlapped slots, a conflict called *slot blocking* happens. Because of the slot blocking and burst collision constraints, the CE could be enabled to attribute a slot to any source-destination pair.

The centralized allocation mechanism can be seen as a resource constrained scheduling problem [103]. Each source can be considered as an independent processor. The sequence of slots to attribute to the source corresponds to a sequence of jobs attributed to a processor and the data cycles related to each destination corresponds to sets of resources. In this model, a processor cannot perform more than a job at a time and each job requires only one resource. The main task of the allocation mechanism is to attribute job to resources by obeying blocking constraint. Hence, as the resource constrained scheduling problem is NP-complete [104], the resource allocation scheme in TWIN is NP-complete too. In the section IV.4, we propose some heuristics to perform this functionality.

In the distributed scheme, a CE is located at each destination node. It manages burst transmission for only sources having traffic to send to this destination. As control entities run their resource allocation algorithm independently, slot blocking can occur at the source side and in this case the source has to choose the convenient destinations to send burst.

In the centralized scheme as well as in the distributed scheme, all data cycles belonging to the same control cycle use the same allocation configuration. This approach allows the reduction of the number of exchanged messages in the control plane.

IV.3.4. Slot assignment mechanism

This functionality is performed at the source side. It consists in attributing bursts to slots. In the distributed scheme, the source has to manage the various grants (coming from various destinations) and chooses the adequate slot to use in the case of slot blocking. In the centralized scheme, because of the absence of slot blocking and its resolution at the CE side, this process is reduced to a simple attribution. In the distributed scheme, we propose an algorithm to perform the choosing of slots that we describe in section IV.4.1.1.

IV.4. Centralized vs distributed control planes

In this section, we propose algorithms for the centralized and distributed control planes and we compare their performance.

IV.4.1. Schemes description

IV.4.1.1. Distributed scheme

In the distributed scheme, each source node *i* sends, at a precise moment during the control cycle, one request message per corresponding destination asking for resources. The CE at the destination node *j* collects all the requests received from the sources and computes the proportion of resources R_{ij} that it will allocate to a source *i* by the following manner:

$$R_{ij} = \frac{r_{ij}}{\sum_{k=1}^{s} r_{kj}} \times n \qquad \textbf{Equation IV-10}$$

r_{ij} represents the number of required slots by the source *i* to transmit traffic to the destination *j*, *s* is the number of sources related to the destination node *j* and *n* represents the number of slots per data cycle. Besides, the CE buffers the normalized requests and serves them one by one as shown in Figure 32. For each request, it allocates the R_{ij} slots randomly. The random allocation in the distributed scheme avoids losing the same attributed slot at each data cycle when performing the slot assignment mechanism.

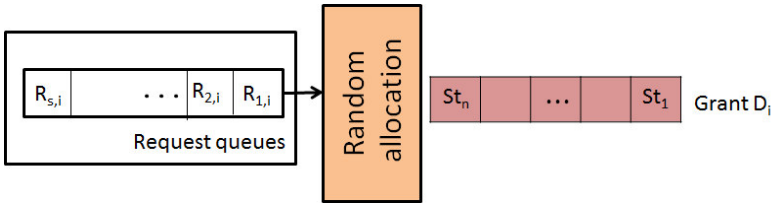


Figure 32- Resource allocation in a distributed control plane

In each control cycle, the source receives many grants (one grant from each CE). As CE executes the allocation mechanism independently of others, the source could have more than one permission to send burst at overlapped slots. Let's assume that source is equipped by one transmitter and at the time *t,* it has more than one attributed slots which are overlapped. Firstly, the source computes the degree of each of the overlapped slots. The degree of a slot is the number of its overlapped slots that are attributed to the same source but intendant to reach

different destinations. Then, the source chooses the slot having the minimum degree *deg*. This means that the source chooses the slot that penalizes the minimum number of flows. If more than one slot has a degree equal to *deg*, the source checks the length of packet queues corresponding to their destinations. The slot corresponding to the destination having the highest queue length will be chosen by the source. At this level, the source gives the priority to the flow having the highest number of waiting packets. In the case of equal queue lengths, the slot corresponding to the destination that has the longest time without being served by the source is chosen. In example of Figure 33, the destination D_1 attributes the first slot to the source; destination D_2 attributes to it the slots 2 and 4, while the destination D_3 attributes to it the slot 5. At time *t*, a blocking slot event occurs between the slots 1, 3 and 5 of D_1, D_2 and D_3 respectively. The slot 1 of D_1 and 3 of D_2 have the same minimum degree (deg=2), however the packet queue corresponding to D_2 is the longest. So, the slot 3 of D_2 is chosen.

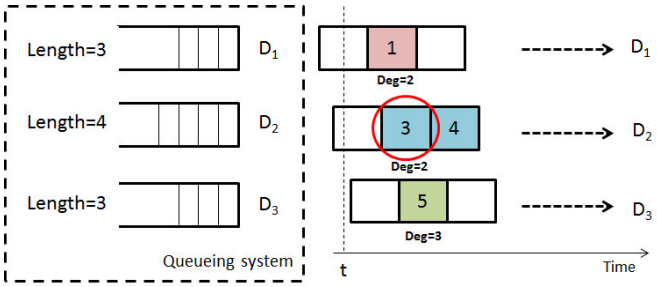


Figure 33- Slot assignment in the case of distributed control plane

IV.4.1.2. Centralized scheme

In the centralized scheme, the CE collects all the flow requests. Then, it determines the number of slots R_{ij} to allocate to each source-destination pair *(i,j)* as follows:

$$R_{ij} = \min(R'_{ij}, R"_{ij}) \qquad \textbf{Equation IV-11}$$

Where,

$$R'_{ij} = \frac{r_{ij}}{\sum_{k=1}^{s} r_{kj}} \times n \qquad \textbf{Equation IV-12}$$

$$R''_{ij} = \frac{r_{ij}}{\sum_{k=1}^{d} r_{ik}} \times n \qquad \textbf{Equation IV-13}$$

r_{ij} represents the number of required slots by the source *i* to transmit traffic to the destination *j*, *s* is the number of sources related to the destination node *j*, *d* is the number of

destinations related to the source node i and n represents the number of slots per data cycle. In this way, the CE normalizes the demanded slots according to the number of available slots per data cycle.

Afterwards, the CE begins the resource allocation process. This process is one of the important tasks of the MAC layer, since it has to manage the bandwidth repartition among the nodes such that it satisfies the maximum of flows. As demonstrated in section IV.3.3, this mechanism is NP-complete in TWIN technology. Therefore, at this step of study, we propose a heuristic approach to perform the resource allocation mechanism.

For this purpose, the CE buffers the normalized requests (R_{ij}) in queues and creates the slot allocation patterns according to a *first-fit* algorithm where, the CE attempts to reserve for a given slot request the first available slot that meets the two following conditions: neither burst collisions occur in the destination node nor slot blockings occur in the source node. The strategy that the CE uses to determinate the order of serving requests has an impact on the obtained resource allocation pattern. We choose to study two different strategies.

In the first strategy, the CE buffers requests according to their destination and treats them successfully as shown in Figure 34. Accordingly, CE accomplishes the reservation by privileging the attribution of *contiguous slots* for the same source-destination pair.

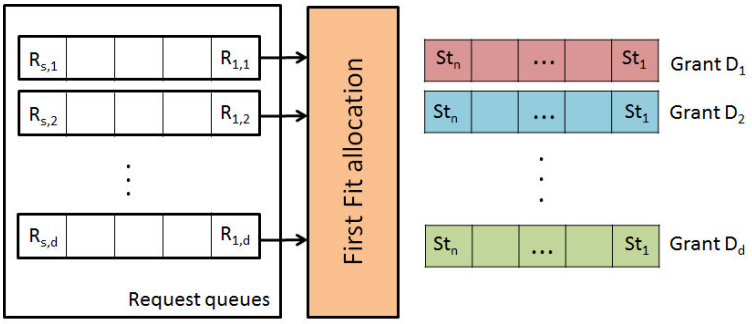


Figure 34- Contiguous resource allocation in the centralized control plane

In the second strategy, the CE treats the requests intended to the same destination in circular order according to a round robin approach as shown in Figure 35. In each round the CE attributes one slot to each flow and it decrease its request by 1. Accordingly, the attributed slots for a given flow are *disjoint*.

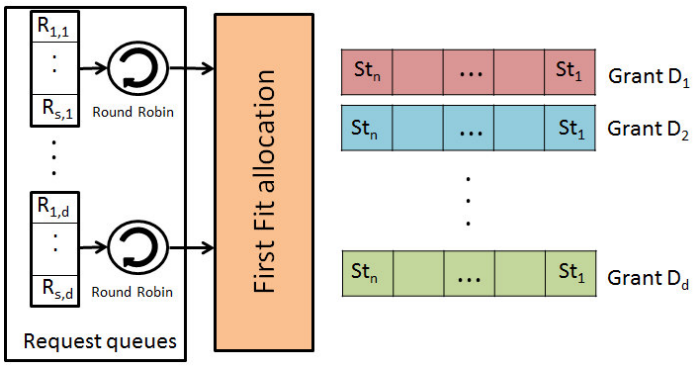


Figure 35- Disjoint resource allocation in the centralized control plane

The two allocation strategies generate two different burst allocation patterns as shown in Figure 36. In the first one, slots attributed to each source to reach a given destination are arranged side-by-side when it is possible. We refer to this strategy as *contiguous allocation*. The second one the attributed slots are scattered throughout the data cycle. We refer to this strategy as *disjoint allocation*.

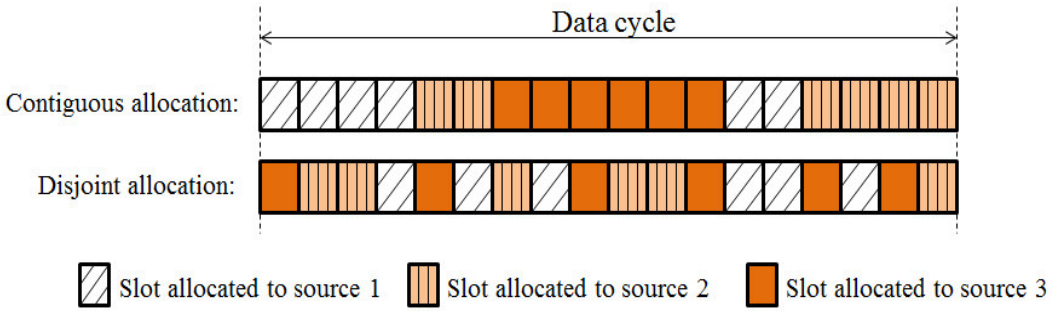


Figure 36- Contiguous and disjoint slot allocation

IV.4.2. Simulation results and discussion

We compare the performance of the proposed control planes using a simulator based on OMNET++ software. Specifically, we consider in this comparative study the distributed control plane and the centralized control plane with either contiguous or disjoint resource allocation. In order to guarantee the reliability of the results, we verify that the confidence intervals are sufficiently small with regard to the system model of this study. Thus, we perform 50 runs for each simulation with the same parameters but different random number seeds. We consider a metropolitan network topology composed of four source nodes and four destination nodes with non-equal propagation times between source-destination pairs. As

shown in **Table 2**, distances between pairs are not multiple of slots, so that, at each source, slots are non-aligned.

	Destination 1	**Destination 2**	**Destination 3**	**Destination 4**
Source 1	222	102	117	213
Source 2	73	90	310	201
Source 3	224	187	76	77
Source 4	110	147	189	36

Table 2- Source-destination distance (km)

Each source node sends traffic to all destination nodes. However, bursts considered in evaluating performance belong to only one source-destination flow. We verify that performances are still similar for the other flows. In this study, we consider that bursts are already assembled and we assume that arrival of the bursts from the burst assembly module follows a Poisson process. This assumption is well justified in [105] and [65]. Bursts are not differentiated with respect to class of service and are supposed to be completely filled.

The capacity of Tx and Rx is set to 10 Gbps. The time slots have a fixed duration equal to the duration of a burst plus a guard time equal to 500 ns, in order to take into account laser tuning time and synchronization accuracy issues. Table 3 summarizes the simulation parameters.

Parameters	**Values**
Capacity of Tx/Rx	10 Gbps
Number of Tx/Rx per node	1
Time slot	5 µs
Guard time	0.5 µs
Size of burst	5600 bytes
Number of slots per data cycle	100
Data cycle duration	500 µs
Control cycle duration	10 ms
Damping factor (K)	20
The speed of the light inside the fiber	5 µs/km

Table 3- Simulation parameters

We focus in this comparative study on four main performance parameters: delay, jitter, queue length and total resource utilization. These parameters are evaluated as a function of the offered load. Hereafter, we mean by offered load, the ratio between the average amount of data (per second) intended to a given destination and the channel capacity between both nodes.

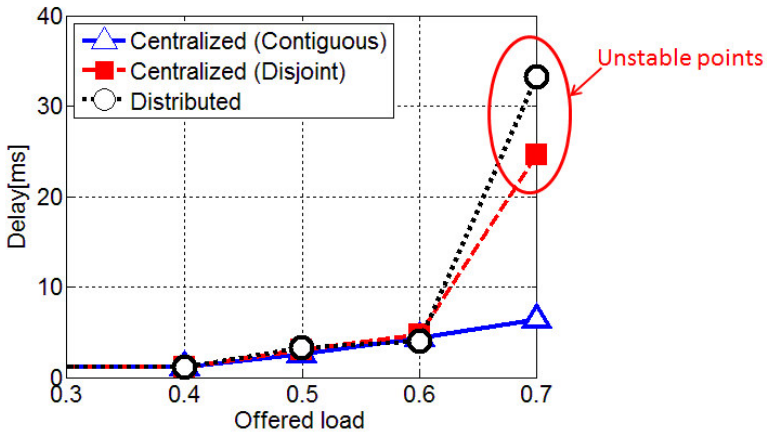


Figure 37- End-to-end delay versus offered load

Figure 37 represents the end-to-end delay seen by the bursts as a function of offered load for the three schemes. The end-to-end delay includes the waiting time, the service time, the transmission time and the propagation delay between source and destination. The waiting time corresponds to the time the burst spends from the entering to the queue until it reaches head of the queue; the service time is the time spent by the burst in the head of the queue waiting for an available slot. The transmission time is the time taken by the transmitter to completely release the burst from the node (it is equal to 5µs). The propagation time between the two studied nodes is equal to 1065 µs (corresponding to 213 km).

We can observe that below an offered load of 0.6 the centralized scheme with contiguous allocation achieves the lowest delay while the distributed scheme performs a delay slightly longer than other schemes. This behavior is probably due to the presence of slot blocking in the distributed scheme which disturbs the burst assignment process. Beyond 0.6 load, the delay for the centralized scheme with disjoint allocation and the distributed scheme increases abruptly (from 4 ms at 0.6 to more than 20 ms at 0.7). The delay in the centralized scheme with contiguous resource allocation increases slowly until a load equal to 0.7 (6.5 ms), besides, it undergoes a sudden rise (38 ms at 0.8).

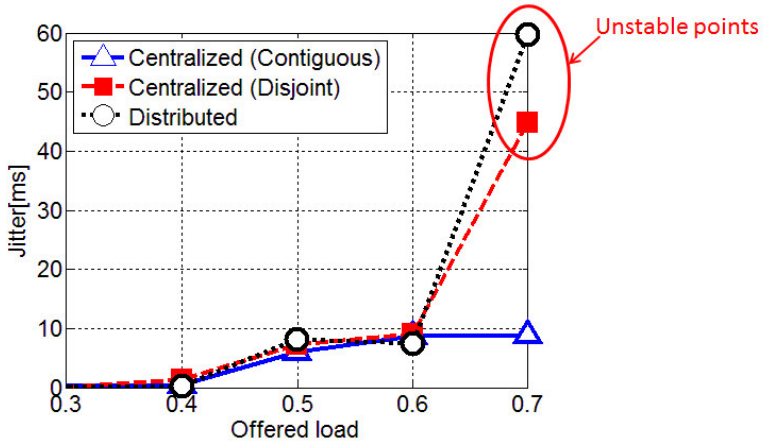


Figure 38- Jitter versus offered load

Figure 38 shows the jitter versus offered load. To calculate this parameter, which represents the variability over time of the latency across the network, we take the difference between the 99th percentile and the 1st percentile of the delay distribution. The jitter curve presents almost the same behavior as the delay curve. For a load between 0.1 and 0.4, the three schemes present a low jitter (1 ms). Then, between 0.4 and 0.6, the jitter increases up to almost 10 ms. Beyond a load of 0.6, the distributed scheme and the centralized scheme with disjoint allocation become unstable (jitter value > 40 ms). The centralized scheme with contiguous allocation shows a steady value of about 10 ms until a load of 0.7.

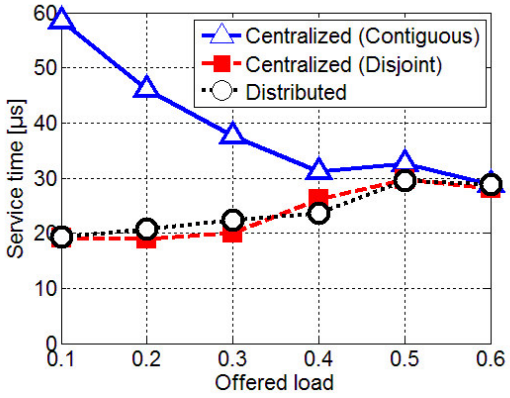


Figure 39- Service time versus offered load

To examine the delay more closely, we depict in Figure 39 the service time versus the offered load. The service time in the centralized scheme with contiguous allocation is higher at low load than at high load: it decreases from 60 µs at a load equal to 0.1 to 30 µs at a load

equal to 0.6. Reversely the service time values in the two other schemes, which are both based on a disjoint allocation, increase from 20 µs at 0.1 up to 30 µs at 0.6 load. This can be explained by the fact that, in the contiguous allocation when a burst arrives to the head of the queue after the end of the block of slots reserved to its transmission, it has to wait for the next data cycle. The higher the load is, the fewer sources loose opportunities to insert bursts. However, in the disjoint case, the arriving burst has more chance to find an available slot in the current data cycle. So, it waits for less time in the head of the queue. In the distributed scheme, the source is still unable to benefit from all its opportunities since it suffers from slot blocking. This becomes more visible at high load (a load superior to 0.5).

According to the previous results, the delay is mainly dominated by the waiting time in the queue.

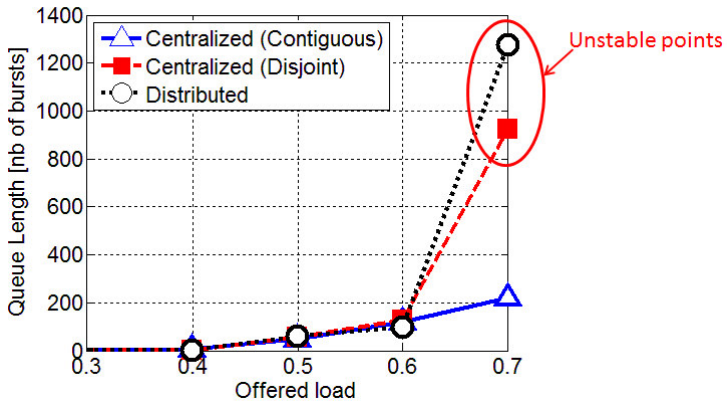


Figure 40- Queue length versus offered load

Figure 40 shows the mean length of the queue as a function of load. In the centralized scheme with contiguous allocation, beyond the load of 0.7, the system becomes unstable and queue size would continue to increase infinitely. For the two other schemes, the system stability threshold is reached earlier at a load of 0.6. In the three schemes, the average of queue length is almost the same for the range of traffic load within which the network is stable. The queue length value explains the significant rise of delay at 0.5 load. In fact, by multiplying the average service time at 0.5 (30 µs) by the length of queue at this load (almost 65 bursts), we obtain a waiting time of about 2 ms, which is consistent with the recorded delay value.

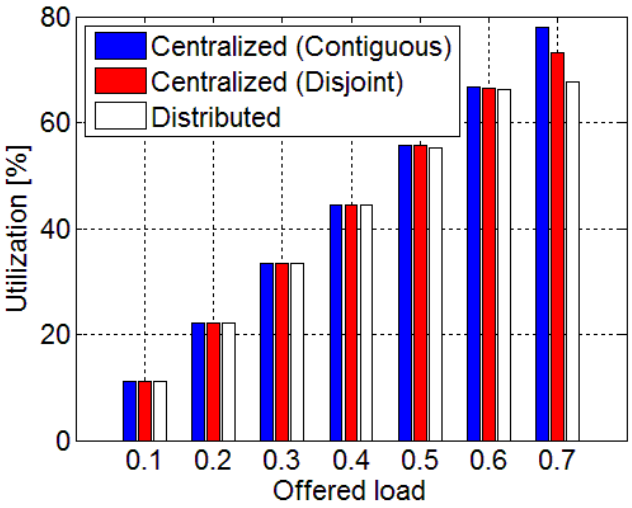


Figure 41- Resource utilization versus offered load

Figure 41 shows the resource utilization (average proportion of used slots during a data cycle) in function of the offered load. The centralized scheme with contiguous allocation presents greater efficient resource utilization than the two other schemes. It can reach a resource utilization ratio of about 80% which enable the emission of more than 7 Gbps of traffic. The percentage of resource utilization in the distributed case is limited to 66%. Compared with the distributed scheme, the centralized scheme with contiguous allocation allows the utilization of almost 20% of additional resources among the available ones.

Referring to the previous results, we conclude that the performance and the stability of the system are related to the ability of control scheme to manage the available resource. The centralized control plane outperforms the distributed one and a deep study of the former approach is worth doing.

IV.5. Centralized control planes

The performance comparison done between the three proposed control schemes (distributed, centralized with two different slot allocation solutions) show that distributed scheme is less efficient than centralized ones. So, in this section, we focus on centralized control schemes. Here, we compare the two aforementioned centralized schemes with a third one based on an optimized resource allocation that we call *static/quasi-static allocation* and a

fourth one presenting the trade-off between an optimized and a heuristic solution. We call it *hybrid allocation*.

IV.5.1. Algorithms description

IV.5.1.1. Static/quasi-static allocation

As the optimal solution of the resource allocation problem in TWIN network is complex, the resolving of this problem cannot be done in real time and the time required to resolve it depends mainly on the number of source-destination nodes in the network. We can assume here that the period of control cycle is enough large to re-compute appropriate optimal allocations taking into consideration the variation of traffic. In this case, we refer to this algorithm as *quasi-static allocation*. In an extreme case, one can suppose that requirements are static and we attribute the slots for each data cycle of a control cycle in a fixed way whatever the traffic variation and whatever the control cycle. Then, the number of required slots to satisfy each flow is calculated once for a fixed traffic matrix. In this case, we refer to the algorithm as *static allocation*.

The allocation mechanism is formulated as an optimization problem. It focuses on maximizing the fill in of grants by taking into account the collision constraints in the destination side, the blocking constraints in the source side and the dimensioning. The dimensioning ensures that each source-destination gets the required number of slot resources.

In this model, we define s as the number of sources, n as the number of slots per data cycle and X_j as a binary vector indicating the pattern related to the reception of bursts at the destination j. The size of X_j is equal to $s.n$. Each index m of the vector X_j could be written as $m = s.(p-1)+i$ where, $1 \leq i \leq s$ and $1 \leq p \leq n$. Each element of the vector $X_{j,m}$ indicates if the slot p is attributed to the source i or not. The purpose of this optimization problem is to find the vector X_j for each destination j.

The optimization problem is modeled as follows:

$$\boldsymbol{max}\left(\sum_{j=1}^{d}\sum_{p=1}^{n}\sum_{i=1}^{s} X_{j,s.(p-1)+i}\right) \qquad \textbf{Equation IV-14}$$

Subject to:

$\forall j, j' \in [1..d]\ \forall i \in [1..s]\ \forall p, p' \in [1..n]$

$$X_{j,s.(p-1)+i} \in \{0,1\} \qquad \textbf{Equation IV-15}$$

$$\sum_{i=1}^{s} X_{j,s.(p-1)+i} \leq 1 \qquad \textbf{Equation IV-16}$$

$$X_{j,s.(p-1)+i} + X_{j\prime,s.(p\prime-1)+i} \leq 1 \qquad \textbf{Equation IV-17}$$

(if slots *p* and *p'* are overlapped at the source *i)*

$$\sum_{p=1}^{n} X_{j,s.(p-1)+i} \quad \leq R_{i,j} \qquad \textbf{Equation IV-18}$$

$R_{i,j}$: is a given, it represents the number of slots to attribute to the source/destination pair *(i,j).*

In the previous model, the constraints in Equation IV-16 avoid the collision in the destination, the constraints in Equation IV-17 avoid the blocking in the sources and the constraints in Equation IV-18 ensure the dimensioning. The constraints in Equation IV-17 require the knowledge of the overlapped slots *p* and *p'* at each source. Therefore, we assume that all data cycles begin at the same time for all the destinations and we consider that the propagation times $\delta_{i,j}$ and $\delta_{i,j\prime}$ between the two source-destination pairs *(i,j)* and *(i,j')* respectively are equal to:

$$\delta_{i,j} = k.\Delta_d + \alpha_{i,j}.\Delta_s \qquad \textbf{Equation IV-19}$$

and

$$\delta_{i,j\prime} = k'.\Delta_d + \alpha_{i,j'}.\Delta_s \qquad \textbf{Equation IV-20}$$

Where, Δ_d is the data cycle duration, Δ_s is the slot duration, $k \in IN$, $k' \in IN$, $\alpha_{i,j} \in IR$ and $0 \leq \alpha_{i,j} < n$. The number of slot offset between these two propagation times at the source *(i)* is equal to:

$$\Delta = \left|\alpha_{i,j\prime} - \alpha_{i,j}\right| \qquad \textbf{Equation IV-21}$$

In a slot aligned scenario, Δ is an integer. If we assume that $\alpha_{i,j\prime} > \alpha_{i,j}$, so the relationship between the two overlapped slots *p* and *p'* intended respectively to destinations *j* and *j'* is expressed by the Equation IV-22:

$$p' = (p + \Delta) mod (n + 1)$$ **Equation IV-22**

Here, "*x mod y*" gives the remainder of division of x by y.

Otherwise, in a non-aligned scenario, Δ is real. In this case, the slot *p* is overlapped with two slots *p'* and *p"*, intended to the destination *j'*. If we assume that $\alpha_{i,j'} > \alpha_{i,j}$, *p'* and *p"* are given by the Equation IV-23.

$$p' = (p + f) mod (n + 1) + 1 \ \ and \ \ p" = (p + f') mod (n + 1) + 1$$ **Equation IV-23**

Where $f = \lceil \Delta \rceil$ and $f' = \lfloor \Delta \rfloor$. $\lceil \Delta \rceil$ is the nearest integer larger than Δ, $\lfloor \Delta \rfloor$ is the nearest integer smaller than Δ.

IV.5.1.2. Hybrid resource allocation

The hybrid resource allocation is based on a trade-off between the static/quasi-static and the contiguous allocation algorithm. Therefore, slots of the data cycle are divided into two parts: fixed part and dynamic part. The resource allocation of slots belonging to the fixed part is done periodically after several control cycles using the static algorithm, while the allocation of the dynamic part is performed in each control cycle using the contiguous resource allocation algorithm. The fixed part and the dynamic part could have the same or different number of slots depending on the requirements of the network.

IV.5.2. Simulation results and discussion

In this study, we compare the performance of the four centralized schemes (disjoint, contiguous, static and hybrid) via the same simulation tool used in the previous study in the section IV.4.2 and we also take the same simulation parameters mentioned in Table 3. In order to guarantee the reliability of results, we verify that the confidence intervals are sufficiently small with regard to the system model of this study. Thus, we perform 25 runs for each simulation with the same parameters but different random seed numbers. The optimization problem in the static allocation scheme is resolved using MATLAB software. Unlike the simulation study in IV.4.2, performance results are computed by taking into account bursts belonging to all the source-destination pairs.

We consider metropolitan network topology composed of four source nodes and four destination nodes with non-equal propagation times between source-destination pairs. We studied two different scenarios:

i) The *slot-aligned scenario* where all the slots are aligned in all sources. Distances between source-destination pairs are mentioned in Table 4.

	Destination 1	**Destination 2**	**Destination 3**	**Destination 4**
Source 1	222	102	117	213
Source 2	73	90	310	201
Source 3	224	187	76	77
Source 4	110	147	190	36

Table 4- Distances in the slot-aligned scenario (km)

ii) The *non-slot-aligned scenario* where no slot is aligned in a source. Table 5 details distances taken for this scenario.

	Destination 1	**Destination 2**	**Destination 3**	**Destination 4**
Source 1	221.90	102.66	117.24	212.82
Source 2	72.70	90.52	310.30	200.68
Source 3	223.64	187.30	76.10	77.10
Source 4	110.10	146.78	189.76	36.54

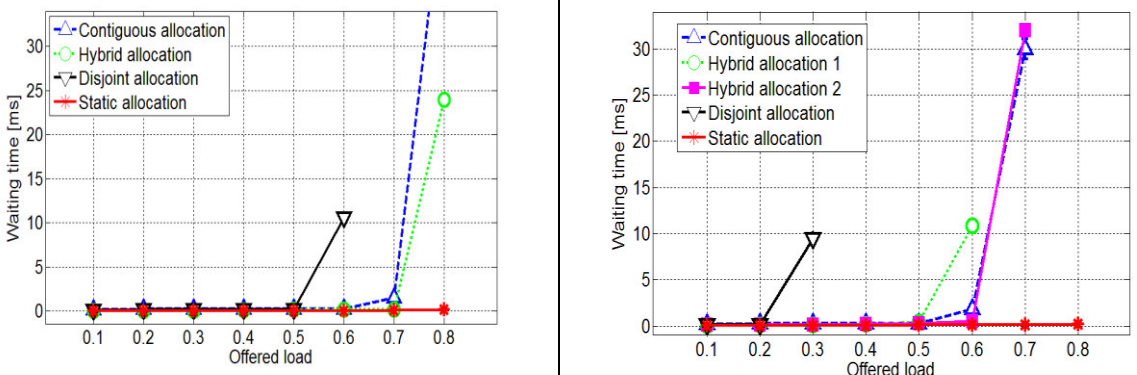
Table 5- Distances in the non-slot aligned scenario (km)

We assume in this comparative study that the centralized scheme based on optimal solution use a static allocation, which means that the allocation of slot does not change during the simulation. For the hybrid allocation, we take the same amount of slots for both the fixed part and the dynamic part (half of the number of slots per data cycle). In the non-slot-aligned scenario, two hybrid schemes are considered. Each scheme uses different slot repartition configuration for the dynamic part. This allows showing the impact of the slot repartition of the fixed part on the system performance especially in the non-slot aligned scenario where the number of overlapped-slots is important.

As in the previous study of IV.4.2, we consider that bursts are already assembled and we assume that arrival of the bursts follows a Poisson process. We focus in this comparative study on the end-to-end burst delay, burst jitter and throughput. These parameters are evaluated as a function of the offered load. As previously mentioned, the end-to-end delay includes the waiting time, the service time, the transmission time and the propagation time. As in TWIN bursts by-pass the intermediate nodes passively without being buffered, the

propagation time depends only on the distance between each source-destination pair. Hence, to show the end-to-end delay, we only present results for the waiting time and the service time.

Curves presented in this section show the performance of each algorithm until the first load at which the system becomes unstable. A system is considered unstable if the length of at least one of its source nodes queues continues to increase infinitely during simulation time.



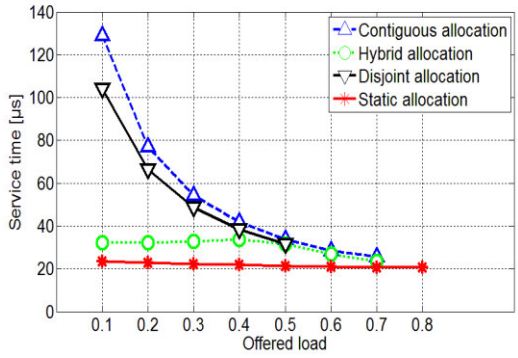
(a) Slot-aligned scenario

(b) Non-slot-aligned scenario

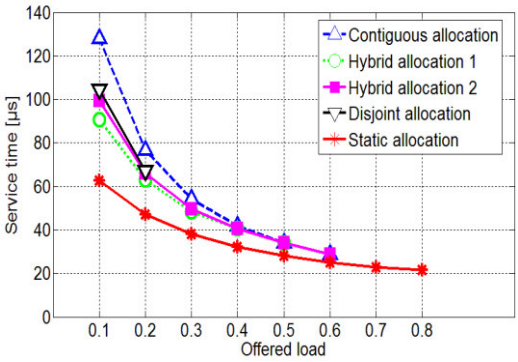
Figure 42- Waiting time versus offered load

Figure 42 represents the burst waiting time in the source side as a function of load. In both slot-aligned and non-slot-aligned scenarios, the static allocation presents the best performance (waiting time of about 0.2 ms up to a load of 0.8) and the disjoint allocation presents the worst one. In the slot-aligned scenario, the hybrid scheme outperforms the contiguous scheme. At a load of 0.7, the waiting time in the hybrid scheme is equal to 0.2 ms while it is equal to 1.5 ms in the contiguous scheme. Both schemes become unstable for a load greater than 0.7.

In the non-slot-aligned scenario, the hybrid allocation is better than the contiguous allocation if we use the configuration 1 and it is worst in the case of configuration 2. This means that, in the hybrid allocation, the choice of the static part influences the performance of the algorithm. It must be done such that it maximizes the opportunities of allocation for the dynamic part.



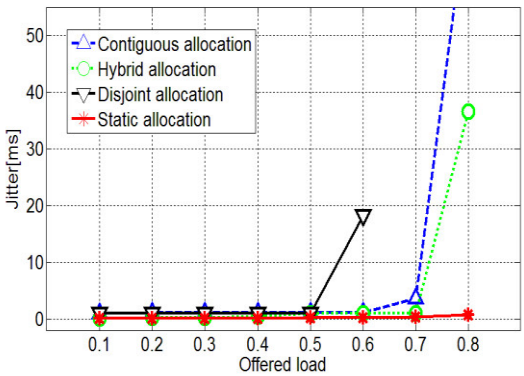
(a) Slot-aligned scenario



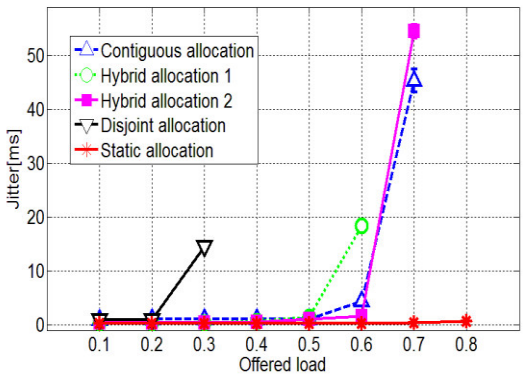
(b) Non-slot-aligned scenario

Figure 43- Service time versus offered load

We depict in Figure 43 the service time versus the offered load. In both slot-aligned and non-slot-aligned scenario, the service time decreases as the load increases. This can be explained by the fact that at low load, when a burst arrives to the system, the queue is almost empty. So, it must remain in the head of the queue for a long time until a slot is available. However, at high load, almost all reserved slots are used and arriving burst remains longer time inside the queue than in the head of the queue. At low load, the mean service time depends on the repartition of the granted slots in the bandwidth and the coincidence between the arrival of a burst and the availability of slots. However, at high load, service times of all algorithms converge to the same value (20µs). This value can be explained by the fact that, in our simulation scenarios, bandwidth dedicated to each destination is divided between four sources. So, on average, a burst at the head of the queue is served after four slots (20µs).



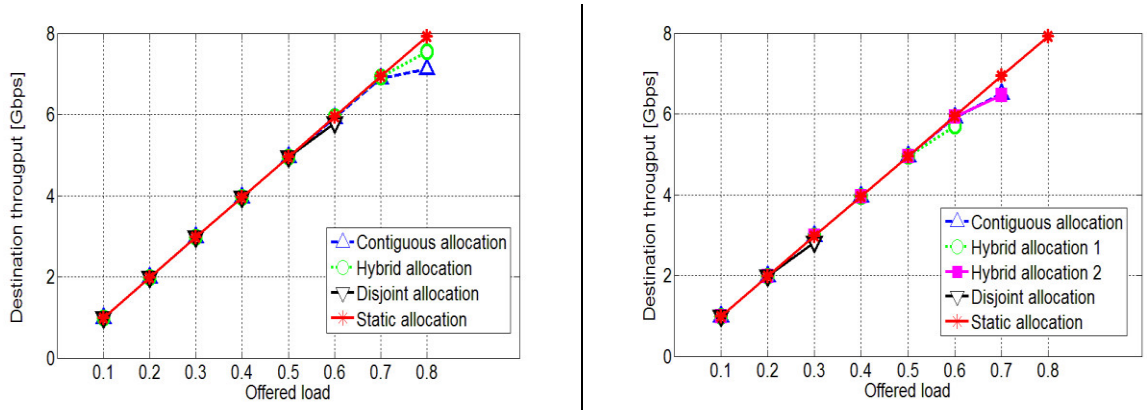
(a) Slot-aligned scenario



(b) Non-slot-aligned scenario

Figure 44- Jitter versus offered load

Figure 44 shows the jitter versus offered load. The static allocation outperforms the other schemes. It compensates its lack of dynamicity by the importance of its bandwidth utilization (the number of slots allocated to each source).



(a) Slot-aligned scenario (b) Non-slot-aligned scenario

Figure 45- Destination throughput versus offered load

Figure 45 shows the destination throughput as a function of the offered load. As the static scheme is based on an optimal slot allocation, it presents the greatest throughput. It can reach a throughput of more than 8 Gbps of traffic. The disjoint algorithm achieves the lowest throughput even in the slot-aligned scenario. This means that its allocation strategy leads to the apparition of a significant number of slots that cannot be attributed to any flow. Consequently, the fact of alternating the allocation of slots between flows within the data cycle using round robin process leads to a bad management of resources. Compared with the slot-aligned scenario, the contiguous and the disjoint schemes loose respectively 14% and 50% of their throughputs in the non-slot-aligned scenario due to the importance of the overlapped slots. For the hybrid scheme, as explained before, the performance is mainly related to the manner of allocating the static part especially in the non-slot-aligned case.

IV.6. Discussion

TWIN concept is interesting in terms of lossless switching and avoidance of optical buffers in the intermediate nodes. It ensures transparency in transit nodes and enables self-routing in the core network because it relies on the wavelength rather than label or address. Nevertheless, the performance of this technology is mainly related to an efficient control plane. In this chapter, we have proposed a new time repartition model and a new general structure for the control plane that are available for both centralized and distributed

approaches. The control repartition model separates the control plane scale from the data plane scale, so that the reactivity of the control plane can be managed by the operator without impacting the data plane time repartition.

As first performance study, we have focused on the comparisons between the centralized and the distributed control planes in terms of end-to-end delay, jitter, queue length and wavelength utilization. Simulation results prove that a centralized scheme with contiguous resource allocation allows a throughput exceeding 7 Gbps. Thus, it outperforms centralized scheme with disjoint allocation by almost 12% and the distributed scheme by almost 15%. In order to better understand the performance of the centralized approach, we carry out a second study focusing on the comparison of four centralized resource allocation schemes (disjoint, contiguous, static and hybrid). The results in terms of waiting time, service time, jitter, and throughput show that, the static scheme performs the best for all parameters as it is optimized for each load. The contiguous scheme achieves an acceptable result with low computational complexity but, it does not guarantee a minimum bandwidth which can be an inconvenient for the prioritized traffic. Accordingly, the hybrid scheme could be a good trade-off provided that the static part is well dimensioned. Results also show that having aligned slot in the source side could improve significantly the performance of the control plane. Theoretically, networks can be designed such that propagation time between each two neighboring nodes is a multiple of a time slot. An example of solution consists in adding Fiber Delay Lines (FDLs) in the output of some nodes. However, in practice, this kind of ideas is not recommended by operators due to the difficulties of maintenance and reparation.

Based on these results, the static/quasi-static centralized approach seems a good candidate for the TWIN control plane despite of its complexity. The complexity of this algorithm is mainly related to the number of nodes in the network. Since we aim primarily the metropolitan area, the number of nodes in the network will not exceed few tens. So, the computational complexity could be overcome by supposing an offline computing and a large control cycle period. Thanks to the proposed time repartition model, the control cycle duration is decoupled from the data cycle duration. So, the increasing of the control cycle period has no impact on the period of data cycle and so, it does not change the number of slots per data cycle. This feature is useful in our static/quasi-static scheme proposal, since considering a

large control cycle period will not increase the number of slots per data cycle. So it will not increase the time complexity of the optimization model.

Despite of its lack of reactivity facing traffic variation, the static scheme outperform the dynamic schemes (disjoint, contiguous and hybrid). However, the performance evaluation has been done assuming Poisson distribution for the burst arrival model which could not give a full view of the behavior of the static/quasi-static scheme facing the real variation of traffic. For this reason, we use in the next chapter real traffic traces in the simulations in order to verify the robustness of this scheme and its ability to manage the abrupt surge in traffic.

Chapitre V. **Packet Level QoS in TWIN**

In this chapter, we propose a new architecture for a metro-backhaul network, called *Multi-hEad sub-wavElength swiTching (MEET)* [10]. Compared with currently rolled out architectures, MEET makes aggregation without several electrical multiplexing stages and replaces them with an all-optical aggregation using a lossless sub-wavelength switching solution based on the *TWIN* concept. According to TWIN, the source nodes are interconnected to each destination node by a multipoint-to-point tree operated on a dedicated *Wavelength Division Multiplexing (WDM)* channel. This concept is used in MEET architecture not only to interconnect the backhaul edge nodes with each other, but also to optically link these edge nodes to different equipment (having separate roles : Internet traffic aggregation, Peering, VoD, PPP sessions...) inside the same POP or to remote core aggregation nodes, outside the backhaul area.

To identify an efficient control plane to MEET, we study the resource allocation strategies presented in the previous chapter in a MEET context. Specifically, we compare a static control plane based on the optimized resource allocation strategy and the dynamic control plane based on the contiguous resource allocation strategy. In the pseudo-static control plane, the resource allocation is formulated as a linear optimization problem, maximizing bandwidth allocation. Since this calculation is a complex process, it is necessary to consider a sufficiently large control cycle duration (from several seconds to several minutes). Hence, the burst emission pattern within the data cycle is kept unchanged for a long period. However, the dynamic or *fast-adaptive* control plane performs the resource allocation for a “short control cycle”. The attribution of slots to flows is done dynamically based on a heuristic approach. In each control cycle, the control plane collects the bandwidth requirements for each source-destination pair. Then, it creates the slot allocation patterns according to a *first-fit* algorithm and distributes them to the sources. This approach is less complex than the first one. Thus, it

can be periodically performed according to a short control cycle duration (several milliseconds).

The transport of data between nodes is ensured by optical bursts built up by assembling electronic packets. The burst assembly mechanism in TWIN is closely related to the data plane performance. In fact, the intermediate nodes operate at full optical capacity without electronic buffering and processing, so they do not introduce any additional delay. We propose two options to perform the burst assembly process. The first option aims to substitute the slotted approach at the data plane side by a timestamp approach, so that the sources benefit from the guard times between consecutive bursts in order to reduce bandwidth waste. From a control plane point of view, this option does not change the computation algorithms at the control entity that continues to perform with the slotted approach to allocate resources. The second option aims to ensure the QoS by giving priority to the flows having specific requirements in terms of latency and jitter. These two options might lead to further improvement of the centralized scheme performances.

Performance evaluation is carried out using a simulation platform fed by real traffic traces captured on Orange's metropolitan network. The QoS delivered to three different classes of service has been assessed in terms of latency and jitter. Obtained results show that a control plane that does not adapt to short-term variations of the real traffic may provide QoS levels compatible with the requirements of an operational metropolitan area network.

V.1. Burst assembly mechanisms

The burst assembler mechanism builds bursts by collecting several packets sent to the same destination. As TWIN is wavelength-based routing solution, the burst does not need a header containing information about the source/destination address or burst size, as it is the case in Ethernet for example. However, other features have to be taken into consideration to ensure the well transmission of the burst.

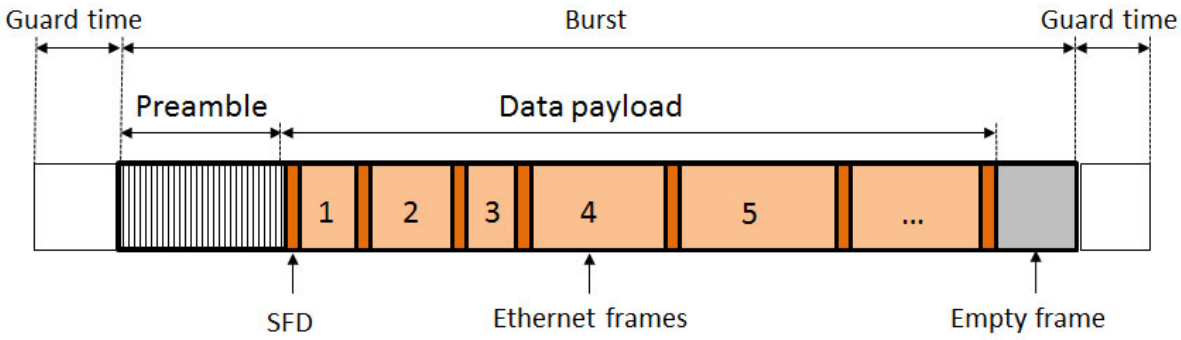


Figure 46- Burst structure

As simple structure of TWIN burst structure, we consider two main parts: the preamble and the data payload as depicted in Figure 46. Preamble does not transport any useful information and is used only for the clock and data recovery of the receiver. It also serves to set the receiver to the appropriate frequency. Once, when a receiver detects the preamble, it starts reading the payload.

Data payload is composed of a sequence of client packets. According to the discussion done in the state of the art chapter, we consider that OBS layer could be a transport layer of some protocols, particularly it has to offer the carrier Ethernet service. Consequently, the data payload could be composed of sequence of Ethernet frames having different lengths. The beginning and the end of each frame is identified using the Start Frame Delineation (SFD). SFD is a unique sequence of bits that is guaranteed not to be seen inside a data frame to avoid the appearance of the delimitation pattern in the data between two real SFD flags. Such false frame delimiter must be modified during the transmission. This could be done by already deployed methods that exist in some protocols such as High-level Data Link Control (HDLC) protocol [106]. Hence, this approach seems like a simplest solution for the burst framing.

The size of a frame could not fit within the remaining space of the current burst. A first approach of this situation consists in the fragmentation of the frame into two smaller frames. The first frame is transmitted in the current burst and the second waits for the next available slot. In this case, a mechanism of frame reassembly should be developed in the destination. The second approach does not allow fragmentation. Hence, the frame is delayed for a later slot and the remaining space in the current burst is filled by an empty frame (stuffing data). This approach leads to under-utilizing allocated slots. We refer to this problem as "*packet*

granularity blocking". To simplify notations and to be coherent with the terminology used in the previous chapter, we will use the term "packet" instead of "frame".

We propose two options for the burst assembly process in TWIN. The first option is related to the management of the available resource and the second option concerns the sensitivity to *ToS*.

V.1.1. Single Slot vs. Multi-Slot assemblers

In the original TWIN concept, a burst is carried in a single slot, yielding per slot overhead due to the guard times. We refer to this approach as *Single Slot-sized burst assembly (SS)*. As an alternative burst assembly mechanism, the source could benefit from the fact that it is fully aware of the future transmission opportunities to build bursts covering several contiguous slots, all assigned to the same destination. In this case, the source manages these contiguous slots as a unique interval of time. This allows building large bursts occupying the transmission time of several slots, which potentially saves some guard times and alleviates the impact of the packet granularity blocking situation. We refer to this new scheme as *Multi-Slot-sized burst assembly (MS)*. In the MS approach, the control entity allocates resources according to a slotted granularity of time and then it sends the grant message containing the indexes of slots to the edge node. The edge sees the contiguous slots as a unique interval of time. So, as shown in Figure 47, the process of merging slots is done at the data plane level and not at the control plane level.

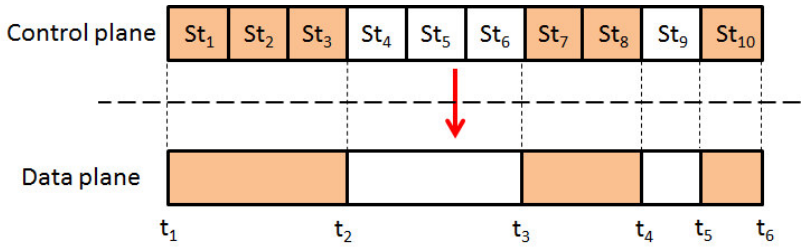


Figure 47- The process of merging slots

Using the MS approach, the assembled bursts have a variable size unlike the SS approach where bursts have a fixed size. Thanks to grants, the source knows in advance the pattern of slot allocation such that it can estimate the size of bursts. If the interval of time dedicated to emit burst is so large and the queue is exhausted, the source has to wait for a specific interval

of time before pursuing the assembly process if the time permits. During this interval, the source could receive new packets and then it can reassemble them into a burst and send them if it has enough time. In our simulation study, this interval of time is equal to one time slot.

V.1.2. ToS-sensitive vs. ToS-insensitive approaches

According to the *"ToS-insensitive"* approach, the incoming packets are firstly classified according to their destinations and are then inserted in a FIFO queue. The assembly process begins little time before the time attributed to emit burst. This burst assembly strategy does not attribute any privilege to packets.

As an alternative burst assembly method, *"ToS-sensitive"* strategy takes into account service priority when building a burst. Packets are buffered in the source node according to their destinations and the value of the Type of Service (ToS) field. The classification of packets into traffic classes relies on the QoS performance objectives in terms of loss, latency, jitter, etc. Here, we consider a three-class model based on the one described in [107]:

- Class 1: real time and interactive traffic, very sensitive to data loss, delay and jitter.
- Class 2: streaming and bulk data traffic, less sensitive to delay and jitter, but still very affected by data loss.
- Class 3: best effort traffic.

When the time attributed to a given destination approaches, the burst assembly is performed according to a Priority Queuing policy, so that highest priority CoS packets are assembled first. This burst assembly method is depicted in Figure 48.

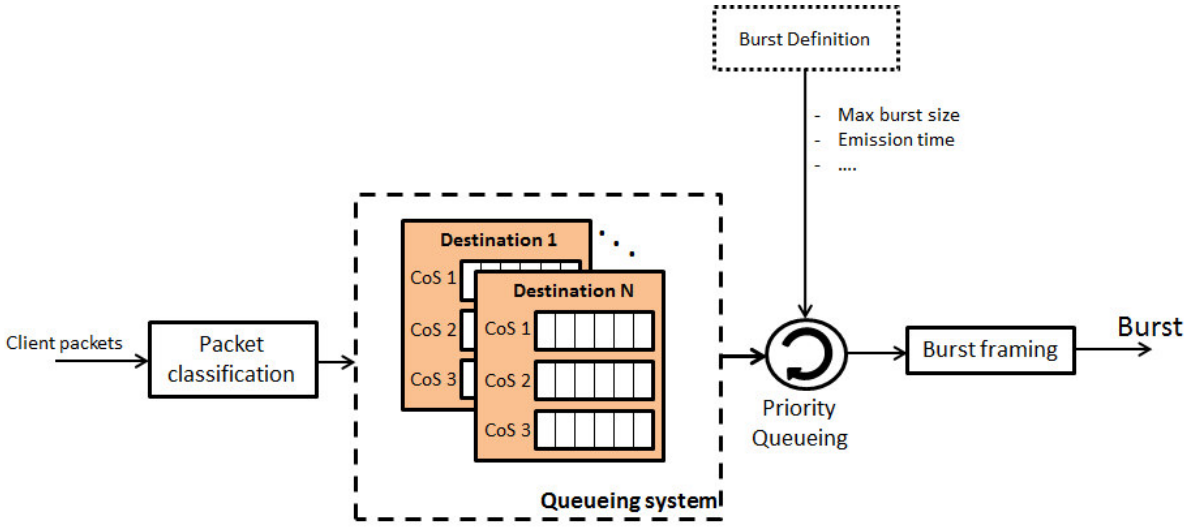


Figure 48- Burst assembly mechanism in the ToS sensitive approach

In both approaches, burst can be composed of packets from different CoS. In an operational network, ToS can be controlled by the network operator (for example, in order to be compliant to multi-class Service Level Agreements (SLA)).

V.2. MEET network architecture

As seen in the second chapter of this report, the current operator networks are designed in a hierarchical way in order to ensure an efficient connectivity. The three levels of hierarchy that are usually defined, namely access level, backhaul level and core level, contain multiple traffic aggregation nodes. A node in a given level aggregates the traffic coming from the immediate lower level, yielding to higher stages of traffic aggregation. At the backhaul level, several access networks are connected to an Edge Node (EN) that, in turn, aggregates traffic and sends it to the Concentration Node (CN). The CN is the first aggregation node in the core network. It is responsible for ensuring connection between the backhaul and the core network. A ring topology is commonly used to link the CN and the EN. In the core network, the CN is connected to different kind of nodes. As mentioned in the second chapter, in Orange architecture, the CN is connected to three main types of core nodes:

- *Regional Nodes (RNs)*: that sends the traffic to higher aggregation levels in the national core network or to other international Tier 1 networks owned by peering partners.

- *Internet Nodes (INs)*: they represent the gateway to the international Tier 1 network owned by the operator.
- *Multiservice Nodes (MNs)*: they permit operator clients to access to the managed service platforms of the operator as Video on Demand (VoD), TV and VoIP services.

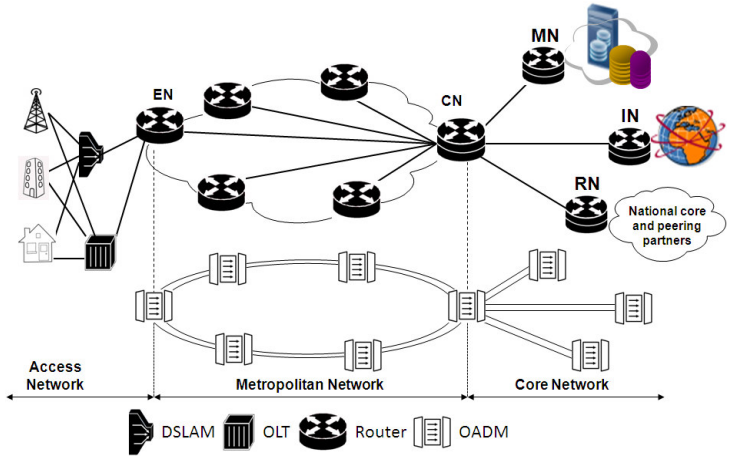


Figure 49- Architecture overview of the current backhaul network

As seen in Figure 49, the current backhaul network has a *"hub and spoke"* structure. Indeed, all the traffic flows are either from the ENs to the CN or from the CN to the ENs. The CN performs an O|E|O conversion to transfer flows between the metro-backhaul and the core network. Therefore, this architecture requires a huge buffering capacity and computing resources in the CN to deal with all the traffic flows.

V.2.1. MEET architecture description

In order to alleviate the traffic load in the CN and provide efficient bandwidth utilization, we propose an alternative architecture, based on the TWIN concept. We refer to this architecture as *MEET architecture*. MEET is TWIN based architecture addressed to the metropolitan network. According to MEET, the metropolitan network is optically extended to reach some core nodes. As an example of application of this solution in the Orange network, MEET enables the EN to be directly connected to the RN, IN and the MN. For this reason, those three nodes are considered as TWIN *remote edge nodes*. They present electronic buffers, they assemble/disassemble bursts and they communicate with the other ENs according to the TWIN control plane. We refer to other nodes as *local edge nodes*. In this architecture, the CN is simply a passive intermediate node. It operates at full optical capacity

without electronic buffering and processing. It represents an optical gateway between the local ENs and the three remote nodes (RN, IN, MN). The MEET architecture is shown in Figure 50. In the current architecture, the communication between local ENs is possible only via the CN, while in this new architecture, they could communicate directly with each other (these connections are not shown in Figure 50 for clearness).

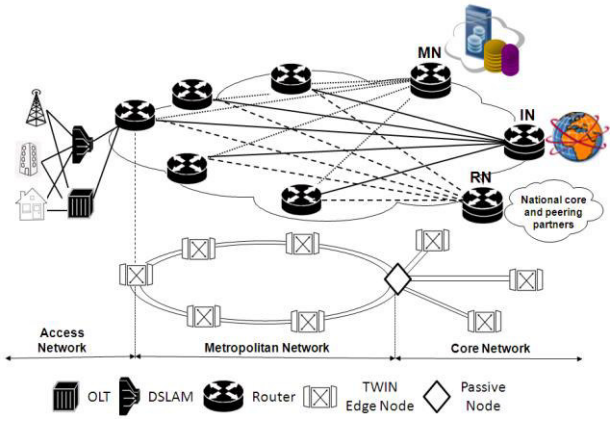


Figure 50- Architecture overview of the MEET

Compared with the current metropolitan architecture, MEET permits an optical aggregation in the CN thanks to the utilization of the sub-lambda technology. Moreover, the adoption of sub-wavelength switching solution could provide both statistical multiplexing and O|E|O interfaces sharing at the edge nodes which enable an efficient use of optical resources. Besides, this architecture is expected to achieve low latency performance compared with the existing one, since it removes an aggregation stage (in the CN), allowing a direct connection between the ENs and the core network nodes. Finally, this architecture provides a more distributed traffic matrix. Indeed, it radically changes the logical metro network architecture from a *hub-and-spoke* to a meshed architecture, which avoids some networking problems like bottlenecks, protection and availability issues at the CN. The physical topology may remain primarily ring-like, but its logical interconnectivity is more meshed.

V.2.2. TWIN control plane for MEET

V.2.2.1. Central point based TWIN architecture

We consider a common start time T of the current control cycle for all destinations. So, the start time T_{ij} of the current control cycle of a destination j at a source i is calculated according to the Equation V-1.

$$T_{ij} = T - \delta_{ij} \qquad \textbf{Equation V-1}$$

δ_{ij} : is the propagation delay between the source i and the destination j.

As shown in Figure 51, slots are not aligned in the source side due to the difference between the propagation delays which could leads to a waste of the bandwidth. This problem is well described in the previous chapter.

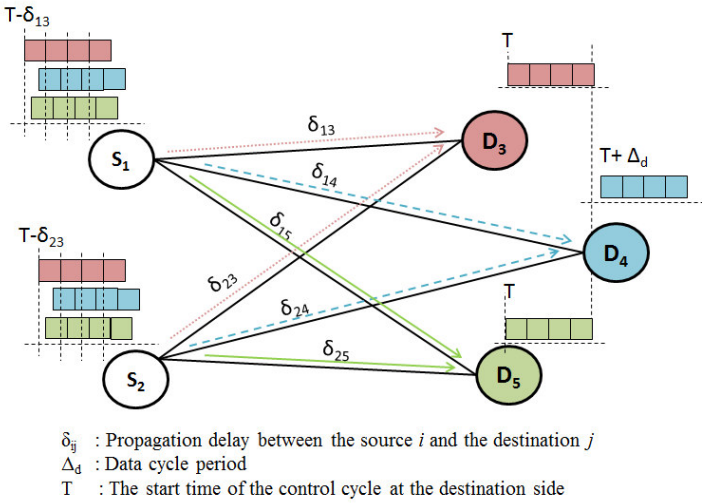


Figure 51- Non-slot-alignment in a non-central point based architecture

As a particular case, we consider a network having a central point W through which all flows pass before reaching the destination. Then, all the optical bursts are merged at the node W before being forwarded to their destinations.

Since there is a single path from the central point to each destination node, avoiding the collision at this point leads to the avoidance of collision at every destination. Hence, the control plane point can consider the central point as a virtual destination for all the flows when computing resources. So, the reference of time could be taken at this point W. In other

words, the beginning time of the current control cycle T is taken according to this intermediate node as depicted in Figure 52.

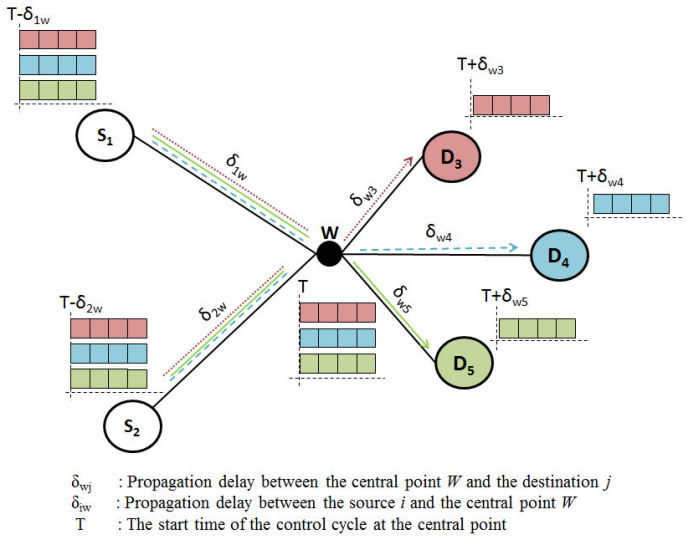


Figure 52- Slot alignment in a central point based architecture

The propagation time between the source i and the destination j, δ_{ij} is expressed by the Equation V-2.

$$\delta_{ij} = \delta_{iw} + \delta_{wj} \qquad \textbf{Equation V-2}$$

Where, δ_{iw} is the propagation delay between the source i and the central point W and δ_{wj} is the propagation delay between the central point W and the destination j. The control plane can consider only δ_{iw} when it computes resource allocations and it takes the beginning time of the control cycles T_i at the source i as expressed in the Equation V-3.

$$T_i = T - \delta_{iw} \qquad \textbf{Equation V-3}$$

Hence, the time of the beginning of the control cycle at the source i is the same for all the destinations j. Consequently, the alignment of slots is ensured even if the distances between the source-destination nodes pairs are not a multiple of slots. At the destination side, the beginning time T_j of the control cycle at the destination j is given by the Equation V-4 and it is different from a destination to another.

$$T_j = T + \delta_{wj} \qquad \textbf{Equation V-4}$$

V.2.2.2. TWIN control plane for MEET

MEET is characterized by the presence of a passive central point that enables an all-optical aggregation of traffic between the backhaul and the core parts of the network. This characteristic makes this architecture compliant, in part, to the central point based TWIN architecture seen in V.2.2.1. However, flows between the local ENs do not necessarily pass through the CN. From a control plane point of view, the central point enables a slot-alignment of the flows between local ENs and the remote EN. However, slots used for the communication between local ENs are not-alignment. To ensure, their alignment, we can design the network in such that paths between local ENs pass also through the central point. This solution is an optional feature in MEET. It could provide a better utilization of the bandwidth at the expense of a possible increase in the propagation delays.

Two main approaches can be considered for the allocation of slots: (i) *pseudo-static* resource allocation, for "long" control cycles (at least a few seconds); and, (ii) *dynamic*, or *fast-adaptive* resource allocation for a "short" control cycle. In the pseudo-static case, the schedule is optimized for a given traffic matrix. Performance degradation, in terms of increased latency and jitter, may occur if the resource pattern, computed on a predicted traffic matrix, cannot accommodate the real traffic offered to MEET. In the dynamic case, the schedule is based on the aforementioned disjoint allocation algorithm where the schedule of emission is recomputed according to the traffic variations observed during the previous cycles.

V.3. Performance study

We compares the respective performance of an optimal schedule obtained for an approximate traffic matrix demand, and a heuristically obtained schedule computed on a more exact assessment of the traffic demands. A heuristic schedule is computed faster than an optimal one, and it is designed to fit with the high dynamicity of real traffic profiles. Nevertheless, as it is heuristically computed, it may thus not optimize the bandwidth utilization. We also evaluate the performance of the different proposed burst assembly options and we compare them.

V.3.1. Simulation framework

In order to assess the efficiency of the proposed mechanisms, we conduct simulation studies using real metro network traffic traces as input. We evaluate the performance of the different control planes in terms of QoS objectives using a simulator based on OMNET++, implementing the MEET architecture. Each node presents a single 10 Gbps transceiver and has infinite capacity queues. Time slot and guard time are respectively equal to 5 µs and 0.5 µs. The pattern considers 100 slots, which yields a data cycle of 500 µs. For the dynamic case, we take a control cycle equal to 10 ms.

The simulated network corresponds to a French backhaul consisting of ten traffic nodes. The distances between the nodes are in the order of a few hundreds of kilometers as depicted in the Table 6. The propagation delay between the furthest node pairs is 1.5 ms, while being lower than 1ms for most of the pairs.

	RN	**IN**	**MN**	**EN1**	**EN2**	**EN3**	**EN4**	**EN5**	**EN6**	**EN7**
RN	0	0	0	136,5	53,8	130,4	153,2	43,2	65,2	32,7
IN	0	0	0	136,5	53,8	130,4	153,2	43,2	65,2	32,7
MN	0	0	0	136,5	53,8	130,4	153,2	43,2	65,2	32,7
EN1	136,5	136,5	136,5	0	191,3	51,2	142,9	93,3	201,7	120,1
EN2	53,8	53,8	53,8	191,3	0	191,1	99,4	97	119	86,5
EN3	130,4	130,4	130,4	51,2	191,1	0	91,7	87,2	195,6	114
EN4	153,2	153,2	153,2	142,9	99,4	91,7	0	178,9	218,4	185,9
EN5	43,2	43,2	43,2	93,3	97	87,2	178,9	0	108,4	26,8
EN6	65,2	65,2	65,2	201,7	119	195,6	218,4	108,4	0	97,9
EN7	32,7	32,7	32,7	120,1	86,5	114	185,9	26,8	97,9	0

Table 6- Distance between couple of nodes (km)

The pseudo-static resource allocation is obtained using CPLEX solver. The simulator is fed by real packet traces corresponding to eight millions packets. The traces have been gathered at peak hour (21:00). The IP snapshot was performed by a probe, placed at the core network border, and equipped with dedicated capture cards able to catch all the packets during the probe process. We thus obtain, for each packet, its source address, destination address, ToS, size and real arrival time. We can derive from this data a set of traffic flows between local ENs and the three remote nodes. We then build artificial packet arrival schedules by multiplying the inter-arrival times by different load factors. This yields realistic traffic profiles with intensities up to 10 Gbps for the most loaded node. The maximal traffic matrix is

illustrated in Table 7. On the basis of this matrix, we deduce less loaded traffic matrices by multiplying it by a load factor ranging from 0.1 to 0.9 (a traffic matrix having a load factor of 1 corresponds to the normalized matrix).

	RN	IN	MN	EN1	EN2	EN3	EN4	EN5	EN6	EN7
RN	0	0	0	2.2	0.7	1.9	2.5	1.1	1.2	0.4
IN	0	0	0	0.2	0.2	0.5	0.4	0.2	0.4	<0.1
MN	0	0	0	<0.1	<0.1	0.1	<0.1	<0.1	<0.1	<0.1
EN1	0.5	<0.1	<0.1	0	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1
EN2	0.2	<0.1	<0.1	<0.1	0	<0.1	<0.1	<0.1	<0.1	<0.1
EN3	0.4	<0.1	<0.1	<0.1	<0.1	0	<0.1	<0.1	<0.1	<0.1
EN4	0.5	<0.1	<0.1	<0.1	<0.1	<0.1	0	<0.1	<0.1	<0.1
EN5	0.2	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	0	<0.1	<0.1
EN6	0.3	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	0	<0.1
EN7	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	0

Table 7- Normalized traffic matrix (Gbps)

In the current hub-and-spoke architecture, all the traffic goes through the CN. The CN is the single head node of the network with the most important traffic load in the upstream and the downstream directions. The traffic in the MEET architecture, as depicted in **Figure 53**, is divided into two parts, the huge amount of traffic passes through the central point to reach the core network (traffic between the local ENs and the remote ENs), while a small part of the traffic remains in the backhaul area (traffic between local ENs). Since the traffic is mostly distributed between the local ENs and the three remote ENs, having a non-slot alignment for the flows between local edge nodes has slight impact on the performance of the allocation mechanism. Hence, in our simulation, we consider direct paths between local ENs.

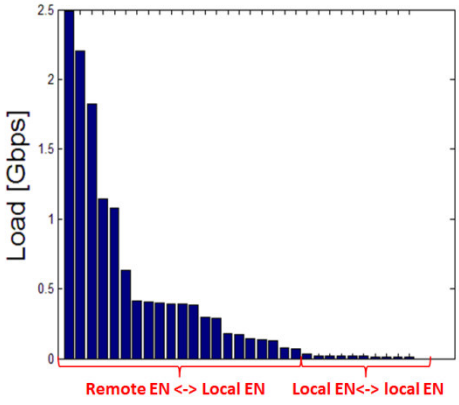


Figure 53- Load of flows in the MEET architecture case

In this study, we focus on the QoS ensured by the different control plane mechanisms and resource management features. We assess whether QoS objectives in terms of latency and jitter meet the values of the Table 8 [107].

QoS	Latency	Jitter	Application
1	3 ms	1 ms	Control, Games, Chat, VoIP
2	5 ms	3 ms	News, E-mail, Streaming, HTTP
3	10 ms	-	P2P, Download

Table 8- Classes of service model

Here, the latency is the sum of the propagation time between the source-destination couple and waiting time which is the time spent by a packet in the source node queue. As TWIN enables a passive optical switching in the intermediate nodes, the main factor of latency is the waiting time, while the propagation time is fixed between each source/destination couple. The jitter is calculated by taking the difference between the 1st percentile and the 99th percentile of the delay distribution. The presented results are for the average waiting time and the jitter of the packets belonging to one of the most loaded flows but it was verified that results concerning the other flows exhibit the same trend.

V.3.2. Traffic dynamicity

In order to understand the dynamicity of traffic, we compare in Figure 54 a real traffic flow presenting a snapshot of the flow between the RN and a single EN, and a theoretical Poisson-based traffic as a function of time. Both traffics are normalized to the same load. We observe that the real traffic fluctuates more than Poisson traffic with instantaneous throughput that could increase from 1.7 Gbps to 3.6 Gbps in only 50 ms.

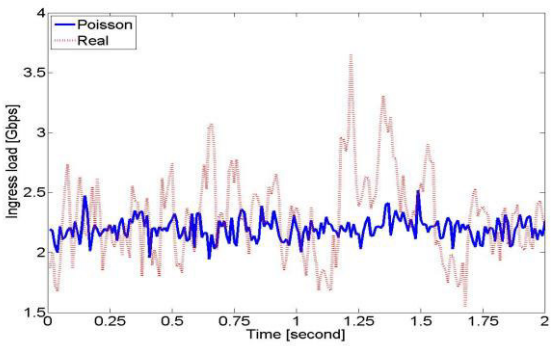


Figure 54- Real and Poisson traffic variations of one traffic flow

We use the same flow and we employ a traffic classification according to the above-mentioned model. Figure 55 shows that the traffic load and the variation are different from a class of service to another. In this particular case, the CoS-2 traffic is the most loaded (72%) and it experiences more dynamicity than the others, while, the CoS-1 traffic is the least loaded.

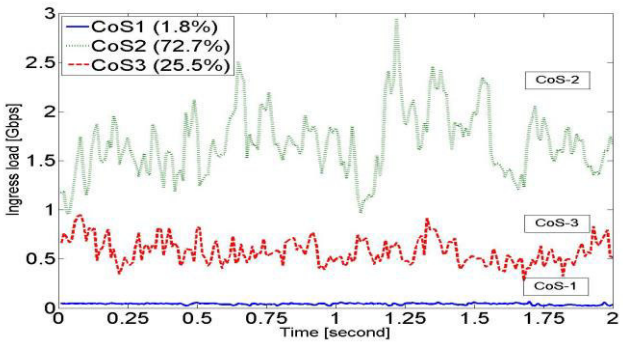


Figure 55- Traffic variations according to the CoS of one traffic flow

V.3.3. Performance evaluation in a ToS-insensitive framework

In this section, we compare the performance of the dynamic and the pseudo-static resource allocation algorithms using the two assembly technique SS and MS with ToS-insensitive approach.

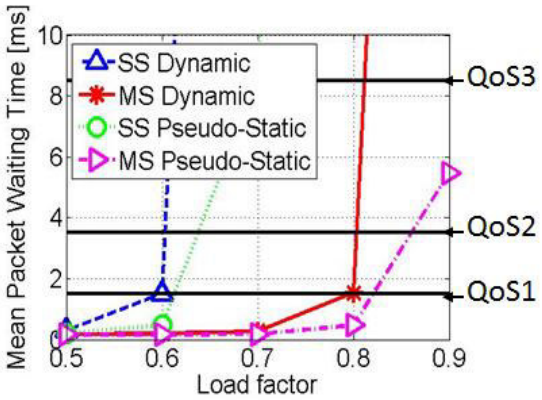


Figure 56- Waiting time in a ToS-insensitive framework

Figure 56 shows the waiting time for all packets. We verified that all ToS classes have the same performance, which is to be expected as packets are served similarly. We first notice that MS results are significantly better than the SS ones which are unable to meet the QoS requirements for a load factor larger than 0.6 for both the dynamic and the pseudo-static

approaches. Using the MS assembly technique, the dynamic and the pseudo static approaches respect the QoS requirements in terms of waiting time are respected until a load factor of 0.8. At load factor equal to 0.9, the pseudo static control plane using MS burst assembly technique respects only the QoS3 objectives.

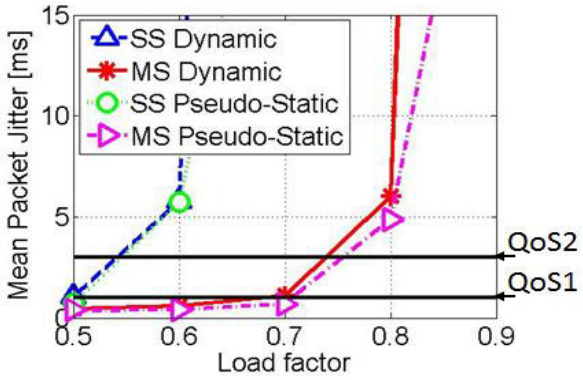


Figure 57- Jitter in a ToS-insensitive framework

Figure 57 shows that the QoS requirements in terms of jitter are met in the case of SS and MS burst assembly methods until a load factor of 0.5 and 0.7 respectively for both control planes (dynamic and pseudo-static). At a load factor equal to 0.7, the MS using pseudo static control plane is slightly more performant than the MS using dynamic control plane.

These two results are first due to the fact that, unlike the SS technique, the MS assembly provides more transmission time since it exploits guard time to send data in the case of two consecutive slots attributed to the same destination. But, this is not the unique reason since the guard time accounts for only 10% of the bandwidth. This is also due to the fact that MS technique alleviates the aforementioned *packet granularity blocking*. Indeed, as a burst in the MS approach is spread over several slots, it is less likely to have packet granularity blocking in the MS approach than in the SS approach, where this blocking is possible in each slot. This is verified in Figure 58, which represents the mean burst lengths for both SS and MS approaches. The SS mean burst size is close to 3.7 µs for all load factors. This is because large packets (1500 bytes) represent a significant fraction of the overall traffic, whereas the time to transmit at 10 Gbps such a packet is large (1.2 µs) compared with the slot duration (4.5 µs). So, the waste of bandwidth due to packet granularity blocking could be alleviated by considering larger slot size (e.g. 10 µs).

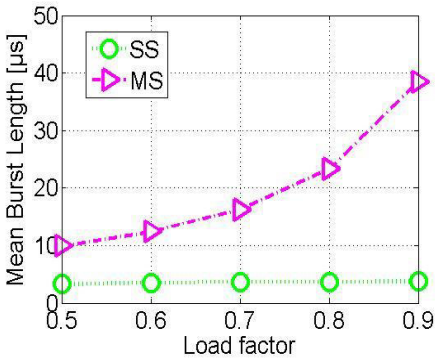


Figure 58- Burst length in the pseudo-static allocation approach.

Results also shows that the MS pseudo-static approach meets latency objectives as long as the load factor is lower than 0.8 but is unable to meet CoS-1 and CoS-2 jitter requirement at this load. It outperforms other control planes including the MS dynamic approach. This can be explained by several factors. First, the pseudo-static approach allows an efficient allocation of resources since this process is based on an exact optimization procedure, which yields a larger number of allocated slots per data cycle than obtained with the first fit heuristic. Sources can thus deal more efficiently with traffic variations. Moreover, traffic dynamicity as illustrated in Figure 55 presents only rather short-term oscillations. Therefore, packets buffered during a peak of traffic will be shortly released, when the traffic decreases, even with a pseudo-static schedule. Lastly, an instantaneous reaction by the dynamic schedule to simultaneous traffic peaks from some flows can lead to starving other flows.

V.3.4. Performance evaluation in a TOS-sensitive framework

We have shown that the ToS-insensitive burst assembly process does not yield a good jitter performance at high loads. In this section, we will improve delivered QoS by considering a ToS-sensitive burst assembler based on a MS technique, for the pseudo-static control plane since it has been shown to out-perform the others. Therefore, we add a ToS-sensitive burst assembler module in each edge node, operating according to a strict priority to the highest ToS packets compared with the others.

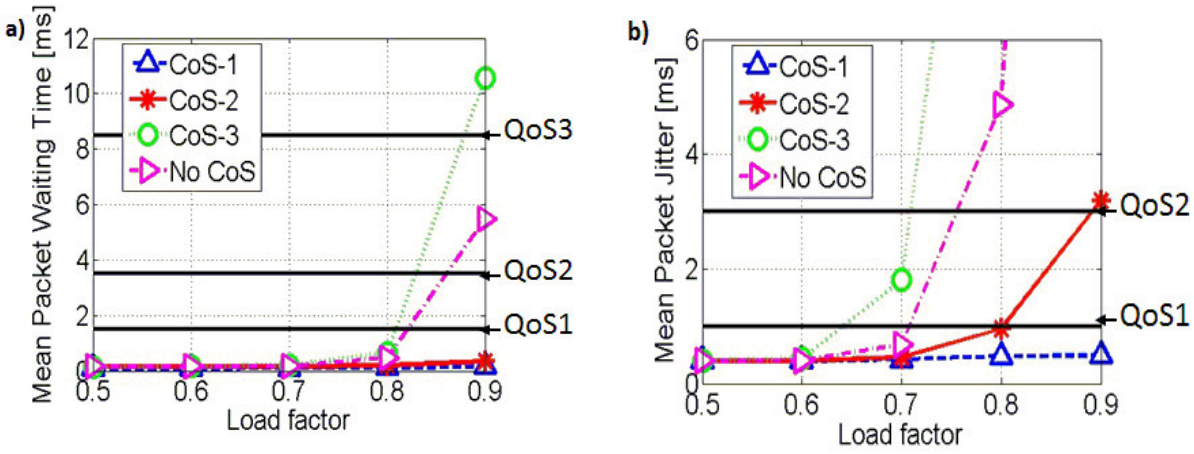


Figure 59- Waiting time (a) and jitter (b) for the pseudo-static-MS control plane

Results in Figure 59 show that ToS differentiation guarantees the QoS requirements for CoS 1 and 2 for a load factor close to 0.9.

CoS-1 packets experience a waiting time lower than 200 µs and a jitter lower than 500 µs. This is not only due to the highest priority of the CoS-1 traffic, but also to its very low load. For instance, Figure 55 shows that CoS-1 traffic for a given source-destination nodes occupies only 1.8% of the total traffic. Therefore, the attributed slots to a given source-destination couple are sufficient to empty CoS-1 queues during a data cycle. This explains well the fact that the waiting time and the jitter remain lower than 500 µs (the data cycle duration).

Despite the high load and the dynamicity of CoS-2 traffic, its waiting time is still less than 1 ms and the jitter is almost equal to 3 ms for a load factor equal to 0.9. This good performance can be explained by the fact that, the CoS-1 traffic is lightly loaded and the CoS-2 has the second highest priority. In fact, in the case of a sudden traffic peak belonging to CoS-2, the assembler attributes few resources to the CoS-1 packets (since they are lightly loaded) and stops assemble CoS-3 packets (since they have the lowest priority) and then CoS-2 packets monopolizes almost all the available resources. As peaks do not last long and pseudo-static plane provides a large bandwidth, the incoming CoS-2 packets are rapidly and efficiently assembled and sent. However, Figure 59 also shows that CoS-3 traffic is significantly penalized, as it receives a QoS worse than the one obtained in a ToS-insensitive framework. This could be alleviated by considering more sophisticated ToS-sensitive frameworks using Weighted Class Based mechanisms instead of Priority Queueing mechanisms.

V.4. Discussion

In this chapter, we have proposed a new architecture called Multi-hEad sub-wavElength swiTching (MEET), that could replace the current electronic backhaul architectures. It is based on a lossless sub-wavelength technology enabling optical aggregation. Thus, it extends the current metro-backhaul architecture by reaching core nodes passively. This allows removing several electrical aggregation stages currently existing between the metro-backhaul and the core networks. Then, it reduces latency and potentially saves energy.

The MEET architecture is optically transparent and support sub-lambda granularity. It presents a good use-case of the TWIN-like operation mode in an operator network. Using simulation and real traffic traces, we have evaluated different mechanisms to implement the control plane for this technology. From the resource allocation point of view, we have compared the performance delivered by a dynamic, fast-adaptive control plane with the one delivered by a pseudo-static control plane. Both packet latency and jitter have been monitored. We have considered different burst assembly techniques (Single Slot-sized burst/Multi Slot-sized burst, Priority based ToS-sensitive/ToS-insensitive).

The results indicate that although there is a significant variation of the real traffic, a pseudo-static control plane with the Multi-Slot approach meets standard QoS objectives of metro-backhaul networks even at highly loaded traffic scenarios.

We have also shown that Single-Slot burst assembly suffers from packet granularity blocking; this could be alleviated by considering longer slots and/or by allowing packet fragmentation, which is a complex process. For this reason, we have proposed Multi-Slot burst assembly to improve resource utilization by alleviating packet granularity blocking and by saving some guard times.

A priority based ToS-sensitive burst assembly process has been shown to deliver excellent performance to time sensitive traffic, by significantly decreasing the delay of non-time sensitive traffic. However, we have to note that the ToS marking of our trace was provided by the application and not overwritten by the network operator, which currently operates a best-effort backhaul. This implies that the ToS field values in the real traffic are not

totally reliable. Moreover, it is to be expected that more sophisticated, weighted class based burst assembly mechanisms would lead to better results; this is to be studied in the future. As future work, we also intend to consider longer traces in order to assess the optimal duration of control cycles.

Chapitre VI. **TWIN Demonstrator**

TWIN is characterized by a simple and passive core node structure but needs a performant control plane to correctly manage bursts emission at the edge nodes. The control plane acts on the source side of the edge nodes via the control unit and it acts on the entire network via the control entity. The control unit at the source side is then the interface between the control entity and the data plane. Its task depends on the kind of control plane. In the case of dynamic control plane, it sends, at each control cycle, a request containing an estimated value of the number of needed optical resources. Besides, it receives a grant containing the new burst transmission pattern to follow during the next control cycle. In the case of static control plane, it only receives grants and transmits them to the data plane. On the other hand, the control entity computes the grants according to the traffic matrix or the request messages received from the control unit at the edge nodes. The computation takes into account the blocking constraints at the source side and the collision constraints at the destination side.

Compared with the dynamic conditions, the static control plane alleviates the reactivity requirements at both the control entity and the control unit. Indeed, grants are computed off-line and do not need to be computed in real time. Thus, the on-line role of the control entity is abbreviated to inform sources about the new burst transmission pattern in each control cycle. Moreover, the control unit does not need to estimate its resource requirements during a short control cycle; even the utilization of a predefined traffic matrix can be considered.

The burst emission unit, at the source side of the edge node, has the role of emitting bursts at the right moment and on the right wavelength according to the pattern provided by the control entity via the control unit. Thus, the error tolerance of the burst transmission time and the wavelength switching duration should not exceed the guard time of several hundred of nanosecond; otherwise, a collision between two bursts sent to the same destination could occur at the intermediate node.

In this chapter, we describe the test-bed that we designed as a Proof of Concept (PoC) for the TWIN solution. The main objective of this test-bed is to experimentally validate hypotheses adopted in the previous simulation study and prove their feasibility. These hypotheses mainly concern the implementation of the control plane and the order of magnitude of some parameters such as guard time, the slot duration, etc. This is not the first time that a test-bed based on TWIN technology is realized. A first demonstrator was done in Shanghai Jiao Tang University in 2007 [108] as it is mentioned in the chapter III of this report. In our test-bed, which is included in the Celtic-Plus project SASER SaveNet [109], we focus on the implementation of a TWIN system based on a static control plane. The test-bed is composed of two source nodes, one intermediate node and two destination nodes as depicted in Figure 60. Compared with the first TWIN test-bed, our demonstrator has more flexible components enable it to support more control plane functionalities and to achieve higher data rate.

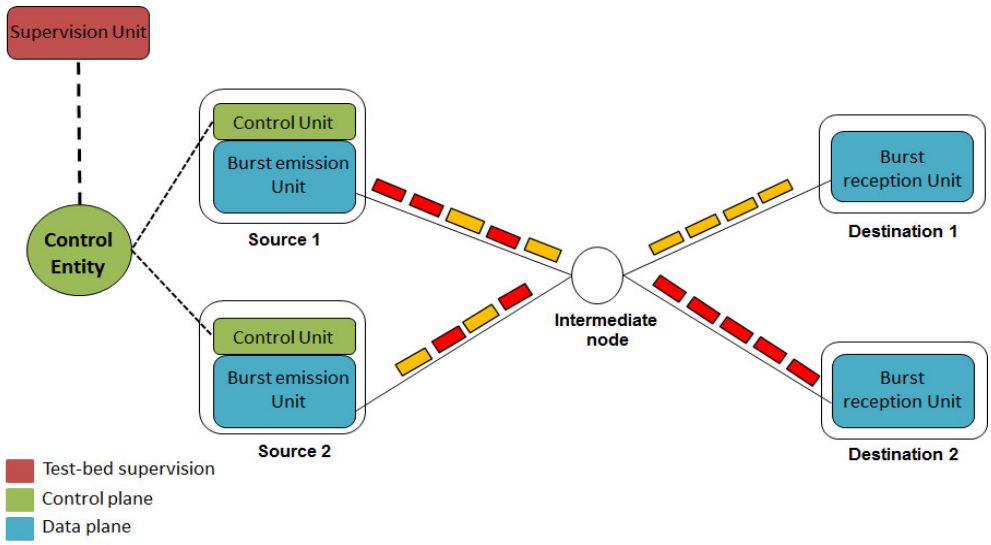


Figure 60- SASER test-bed architecture overview

By knowing the burst transmission patterns in advance, the control entity sends grant messages (patterns) to the control units for each control cycle. The control unit, in turn, transfers these patterns to the burst emission unit that operates at a bit rate of 10 Gbps. The control entity is an electrical system managed by a real-time module that guarantees the response within strict time constraints. The data plane, composed of the burst emission unit and the burst reception unit, is an optoelectronic system managed by Field-Programmable

Gate Array (FPGA) modules. All the system is monitored by a supervision unit allowing the turn on/off and the configuration of the overall test-bed. In addition to these main modules, we use additional components to perform voltage settings, synchronization between modules and measurements.

This test-bed requires robust coding and configuration tools which ensure both the interconnection between the different elements and the implementation of a TWIN control plane. For this purpose, we choose National Instrument (NI) devices to build the control plane. These devices are known by their high performance and rely on LabVIEW framework that provides a powerful and a graphical programming environment to control instruments, to generate and acquire signals and to design embedded systems.

In this chapter, we give, firstly, a general overview of the architecture of the test-bed. Besides, we describe each component by highlighting its characteristics and its internal operation. Then, we present a functional description of the test-bed where we focus on the implemented mechanisms and the interaction between the different components. Finally, due to the fact that some components are not yet implemented, we show the obtained results for only the control plane and the burst emission unit. The experimental results are print-screens extracted from a spectrum analyzer and two oscilloscopes.

VI.1. Organic description

A part of the test-bed components is shown in Figure 61. The supervisor unit (1) enables the user to monitor and configure several devices. The control plane platform is supported by a chassis containing a real-time controller (2) representing the control entity and an FPGA card (3) representing the control unit at the source side. The FPGA is also a key element of the burst emission unit since it also monitors the components that ensure this process. Therefore, it is coupled to an adapter device (4) that generates signals with the required voltage levels. The output signals provided by the couple FPGA/adapter pass through a second adapter (5) equipped with SMB connectors in order to have an easy way to connect the FPGA to the other devices. The optical signal generated by the tunable laser (6) is modulated by an external modulator (7) according to the electrical pattern coming from the Burst Pattern Generator (BPG) (8). An external clock generator (9) feeds the FPGA and the

BPG by the adequate clock rate in order to ensure the frequency synchronization between them.

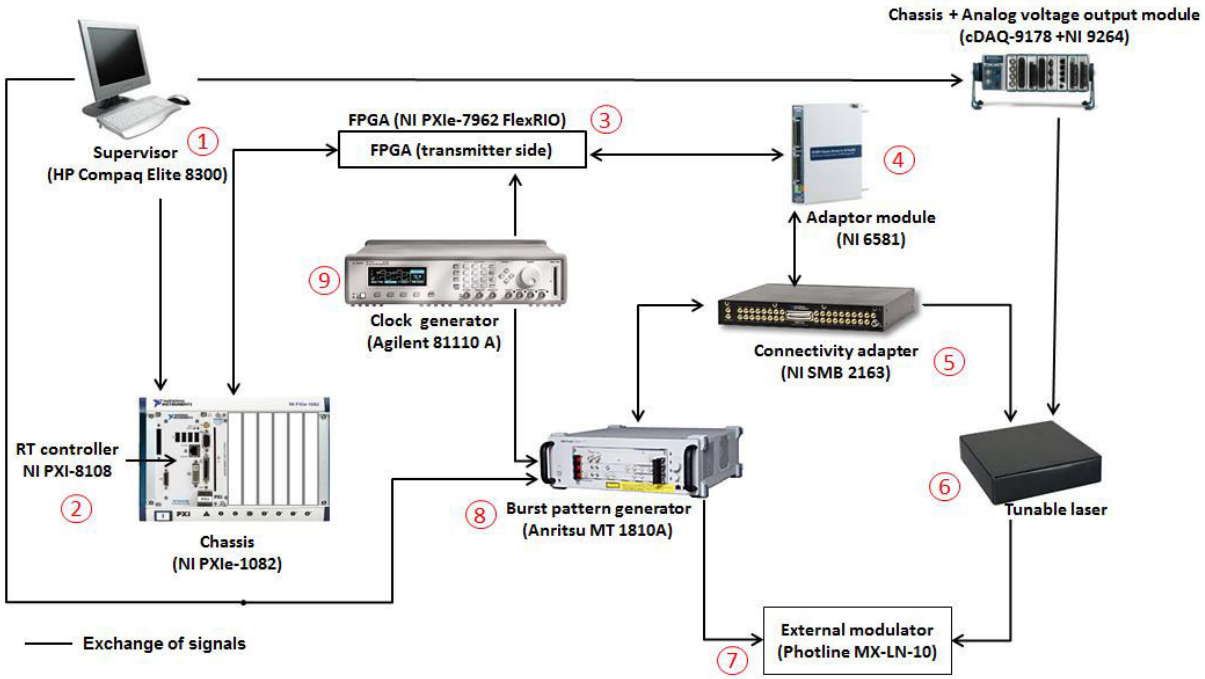


Figure 61- Overview of the test-bed components

In the following sections, we will describe these test-bed components in details. For the presentation, we group them into four different classes: supervision and control entity components, burst emission unit components, synchronization components and intermediate node components.

VI.1.1. Supervision and control entities components

VI.1.1.1. Supervision unit

The test-bed is monitored by a Hewlett-Packard supervisor computer [110] (component (1) in the Figure 61). The computer has an Intel Pentium CPU 2.90 GHz processor and 4Gb of RAM and it is operated by Windows 7 Enterprise. In order to ensure its monitoring role, the supervisor computer is equipped with two Ethernet cards relating it to two local networks. The first local network connects the supervisor computer to the control plane platform and the internet. The second network connects the supervisor computer to the BPG in order to manage

the electrical burst parameters. These two local networks based architecture is imposed by the BPG which requires a dedicated local network and a fixed IP address.

The supervisor computer provides a user interface to handle experiments (switch on/off, set the control cycle duration …) and to configurate the components (voltage levels, …). We distinguish three main roles to the supervisor. Firstly, it ensures the communication and the interconnection with the control entity. Secondly, it provides an interface to configure the BPG by defining the payload and length of bursts, the throughput and the physical characteristics of the output and input signals (voltages). Finaly, it has another interface with the tunable laser (via a Labview program) enabling the setting of the wavelength addresses and the adjustment of the required values of the currents on the different sections of the fast tunable laser to get the desired wavelengths.

VI.1.1.2. Control entity

The chassis supporting the control plane framework of the test-bed is a NI product having as reference NI PXIe-1082 [111]. It provides eight slots that support all the components for the test-bed control plane. It features a high-bandwidth backplane to meet a wide variety of high-performance tests, measurement and control application needs. The chassis also incorporates timing and synchronization features, including built-in 10 MHz and 100 MHz reference clocks with an accuracy of ±25 ppm (parts per million).

Combining the PXIe-1082 chassis with a compatible embedded controller results in a fully compact computer (component (2) in the Figure 61). The embedded controller should occupy the first slot of the chassis in order to have the correct connectivity. We chose the NI PXI-8108 Compact PCI [112] as embedded controller for its high-performances: it has an Intel Core 2 Duo 2.53 GHz processor, 80 GB hard drive and 64-bit DDR2 socket that can hold up to 4 GB. This PXI embedded controller can be configured to boot into a real-time operating system in order to carry out a deterministic and accurate events, required by the control plane of the test-bed. The implementation of the control entity's algorithms is done via a graphical real-time programming language provided by LabVIEW.

VI.1.2. Burst emission unit components

VI.1.2.1. FPGA Module

As aforementioned, the FPGA (component (3) in the Figure 61) represents the control unit at the source side since it receives the grant and it is also a part of the burst emission unit since it configures the transmitter devices. To ensure compatibility with the already chosen hardware (PXIe-1082 and PXI-8108) and the adequate Input/Output (I/O) signal speed (up to 100MHz), we opt to an NI FPGA card: NI PXIe-7962R [113]. NI PXIe-7962R belongs to FlexRIO family of NI. This FPGA module features Xilinx Virtex-5 SXT FPGA with 512 MB of onboard DRAM and 4.752 Kbits of FPGA memory. The FPGA device is plugged in the PXI chassis and it is featured by a specific integrated circuit to provide a high-bandwidth communications link with the backplane of the chassis. The Virtex-5 FPGA architecture is optimized to efficiently use computing resources and then execute instructions quickly.

This device, programmed with NI LabVIEW FPGA Module, provides high performance I/O and user-defined hardware processing on the PXI platform. In our implementation, we use an interesting feature provided by LabVIEW FPGA, which is the *single-cycle timed loop* (SCTL). This feature optimizes the code and allows the execution of many operations per clock cycle. Figure 62 shows a print-screen of a part of the block diagram of the FPGA program realized with Labview. It contains a SCTL supplied by an external clock.

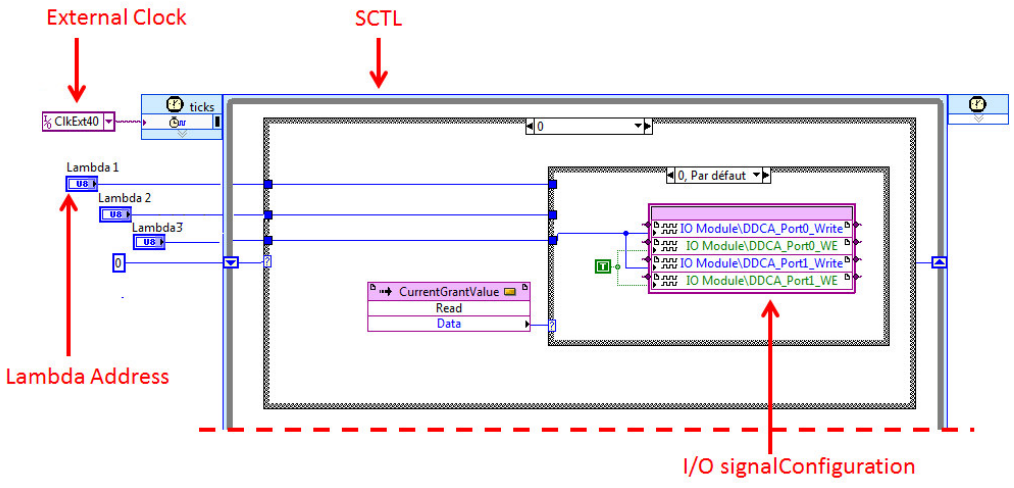


Figure 62- Part of the LabVIEW block diagram of the FPGA program

VI.1.2.2. Adapter Module

The NI FlexRIO FPGA module is coupled to an I/O adapter module in order to manage the digital I/O signals. The adapter module (component (4) in the Figure 61) performs the high-speed communication between the FPGA and the other burst mode transmitter devices. It provides the digital I/O signals as configured by the FPGA. We opt to the NI 6581 adapter module [114] that is compatible with NI PXIe-7962R.This adapter module is able to manage 100 MHz digital I/O. It features 54 single-ended digital I/O lines with software-selectable voltages of 1.8, 2.5, and 3.3V. This configuration fits well with our requirements. The adapter module provides also an input terminal for the external clock acquisition. This feature, as will be explained in the next section, is useful to ensure the synchronization between the FPGA and the BPG modules.

To simplify the connection with other devices, the NI 6581 adapter is connected to a terminal block (component (5) in the Figure 61), providing SMB connectivity (NI SMB-2163 [115]), using a shielded single-ended cable.

VI.1.2.3. Burst Pattern Generator (BPG)

The electrical bursts are generated by a pulse pattern generator that enables high-speed packet transmission. In our test-bed, we use the MT1810A Anritsu chassis [116] (component (8) in the Figure 61). This chassis can support up to 4 plug-in pulse pattern generator or error detector modules. The pulse pattern generator module that we use is the MU181020A of Anritsu [117]. It can generate a variety of patterns including Pseudo-Random Binary Sequence (PRBS) and predefined data with a bit rate that can reach 12.5 Gbpss. It is controlled by the MX180000A Signal Quality Analyzer Control Software installed in the external supervisor computer. The connection between the supervisor computer and the MT1810A is ensured by a dedicated local area interface with a fixed IP address.

VI.1.2.4. Tunable laser

In the source side of the node, we need a tunable laser enabling to quickly switch from a wavelength to another during the guard time. In our test-bed, we use a Finisar S7500 tunable laser [118]. It integrates a Semiconductor Optical Amplifier (SOA) and a tunable Modulated

Grating-Y (MG-Y) laser [119]. The SOA facilitates flexible control of the output power and acts as a shutter when reverse biased which enables dark tuning between channels. The MG-Y laser is an electronically tuned device that can address any wavelength in the C-band. Since no mechanical or thermal adjustments are necessary, channel switching is very fast.

As shown in Figure 63, the MG-Y laser consists of five main sections [120]. The first section is the *gain section* that amplifies the light. The amplified light passes then through the common *phase section* which performs the alignment between the cavity mode and the reflected peaks. The *MultiMode Interference (MMI) section* splits the light into two equal beams. Each beam crosses the *bend section* in order to increase the separation between the waveguides. Besides, each beam goes through the *reflector section* that filters out certain frequencies.

The selection of one lasing frequency is based on additive Vernier effect. A large reflection occurs at the frequency where a reflectivity peak from the left reflector is aligned with a reflectivity peak from the right reflector. The laser will thus emit light at a frequency closest to the peak of the aggregate reflection.

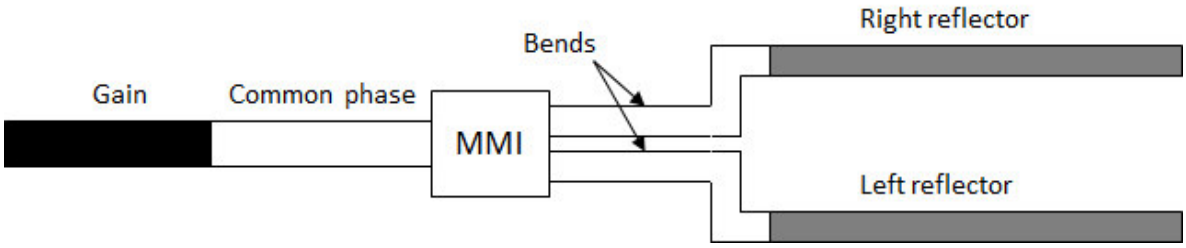


Figure 63- Structure of the modulated grating Y laser

The tunable laser is monitored by five currents: the laser gain current (I_{gain}), the SOA current (I_{SOA}), the reflector currents (I_{right} and I_{left}) and the phase current (I_{phase}). The gain current is kept unchanged during the laser operation. The SOA current monitors the turn on/off of the output optical signal. The two reflector currents and the phase current are carefully chosen to precisely select wavelengths amongst the 89 available wavelengths (50 GHz spaced) in the C-band.

The gain current is delivered by an OptoSCI LDR250GAS board which performs also the temperature control. The reflectors, SOA and phase currents are delivered to the laser by a

specific board (developed by a partner in SASER project) which makes the fast switching between the various groups of four currents corresponding to the different wavelengths.

The fast tuning of the laser from one wavelength to another is managed by an external monitoring device via two signals: the wavelength address signal and the Tx-Enable (Tx-E) signal. The wavelength address is digitized on 8 bits but only 3 are differentiating in our experiment as the specific board is able to manage up to 8 wavelengths. According to the requested wavelength address, the three currents (phase, left and right reflectors) are modified. The Tx-E signal requests turning on/off the emitted light which is performed by varying the SOA current.

The Figure 64 gives the user interface of the laser configuration board. It offers many features enabling for instance to manually turn on/off the laser or to specify the values of the five aforementioned currents needed for each wavelength.

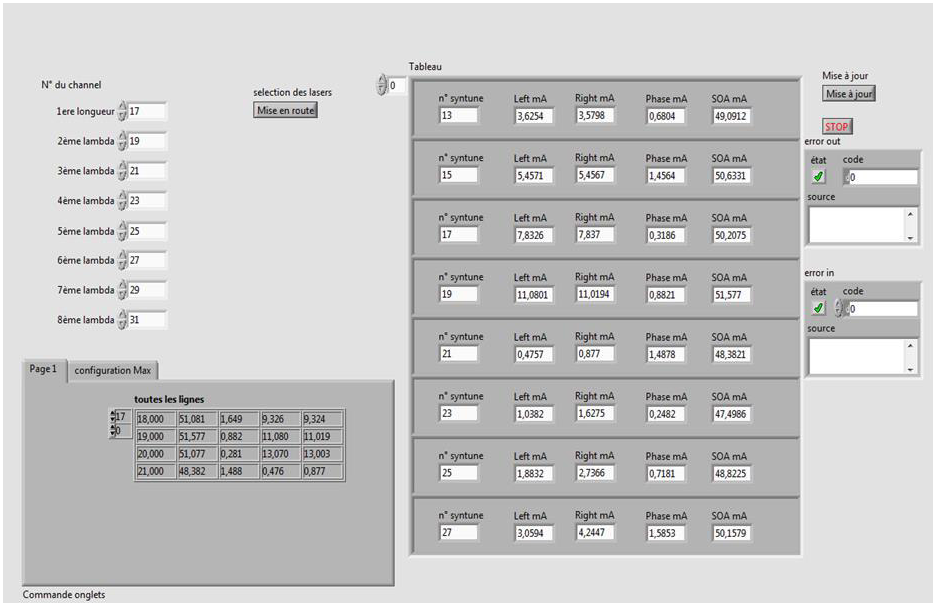


Figure 64- User interface of the laser setting board

VI.1.2.5. The modulator

The role of the modulator (component (7) of the Figure 61) is to print the electrical data generated by the BPG on the optical signal. This can be done either internally within the laser structure or externally using an external modulator. In the internal approach, the modulation of the optical signal is achieved by controlling the current injected into the laser. In the

external approach, the modulator is placed after the light source which emits continuously. In our test-bed, we rely on the external approach as it is the easier way to manage both the modulation of the optical signal with the data at 10 Gbps and the modulation of the wavelength at the rhythm of the bursts. Some work reports the direct modulation of the MG-Y tunable laser at 2.5 Gbps [121] and another at 10 Gbps [122] using the Chirped Managed Laser technology.

In the test-bed, we use a Photline MX-LN modulator based on Mach-Zehnder structure designed for optical communications at data rates up to 12.5 Gbps [123]. The continuous input optical signal is split into two halves. Each half passes through electrically actuated phase controllers, made with lithium niobate (LiNb03) [124]. Then, the two halves are recombined. By properly controlling the voltage levels on the phase controllers (i.e. by injecting a correct voltage on the RF (Radio Frequency) input of the modulator), constructive or destructive interferences occur inside the modulator which results into a presence or an absence of optical signal at the output of the modulator. We have chosen to use a Non Return-to-Zero On-Off-Keying (NRZ-OOK) modulation format which means that a "1" symbol of the binary flow is coded with a presence of optical signal during the whole bit duration (100 ps) and a "0" is coded with no signal during the bit at the output of the modulator. In order to adapt the voltage level of the binary flow coming from the burst pattern generator to the required voltage level at the input of the modulator, we need to amplify the signal delivered by the BPG using an RF driver [125].

VI.1.2.6. Synchronization component

The synchronization is ensured by an external clock generator (component (9) in the Figure 61). It provides a clock reference enabling the frequency synchronization between the FPGA module (PXIe-7962R) and the BPG module (MU 181020A). This solution was chosen after multiple failed attempts to synchronize the BPG directly with the FPGA clock and vice versa. The incompatibility between the clocks signals forced us to use an external clock generator.

The used external clock generator is an Agilent 81110A Pulse Pattern Generator [126]. It provides two synchronous output channels. The first output channel generates a 40 MHZ

clock signal and it is connected to the FPGA module via the *global clock* pin of the adapter module. Whereas, the second output channel generates a 160 MHZ clock signal and it is connected to the burst pattern generator module via the *external clock input* connector.

VI.1.3. Intermediate node components

The intermediate node has a simple structure. It is equipped with two demultiplexers and two couplers as shown in Figure 65. The demultiplexer separates the optical signal coming from the source into two wavelengths (λ_1, λ_2) and forward each of them to the corresponding coupler that combines the signals intended to the same destination.

The demultiplexer that we use is based on a classical approach which is the Bulk Grating Technology (BGT). This method uses a combination of individual micro-optical elements arranged in free-space architecture. The grating is the key element of the architecture. It is a diffractive element that enables angular separations of the wavelengths. The multiplexer also uses lenses and prisms to couple light into the fibers. The demultiplexer has 8 channels spaced by 100 GHz. It uses a flat top configuration to have fairly constant loss along the width of the filters. The insertion loss is inferior to 6db, the Polarization Dependant Loss (PDL) is inferior to 0.1 dB and the -1 dB bandwidth is in the range of 50 GHz.

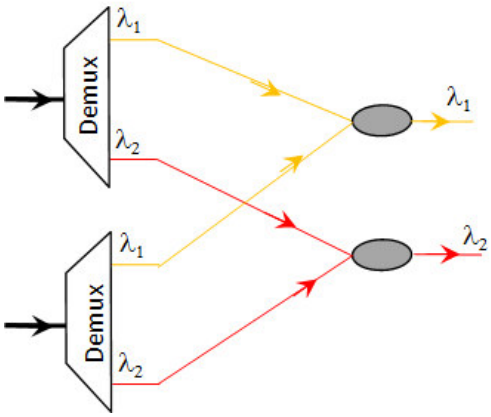


Figure 65- Intermediate node structure

VI.2. Functional description

In this section, we describe the mechanisms and the interaction between the different components that we have already presented in the previous section. We distinguish six types

of signals exchanged between the various components of the test-bed according to their task. As shown in Figure 66, the supervision information corresponds to the information sent by the supervisor, while the control messages correspond to the exchange between the control entity and the control unit (grant message). The command signals represent all signals used to configure the tunable laser and the BPG. The synchronization signals are generated by the clock generator to synchronize the FPGA and the BPG. Finally, the electrical data and the optical signal correspond to the signals generated by the BPG and the laser respectively. The exchange processes are done at different time scales and a perfect synchronization between components is needed to ensure the smooth running of the system.

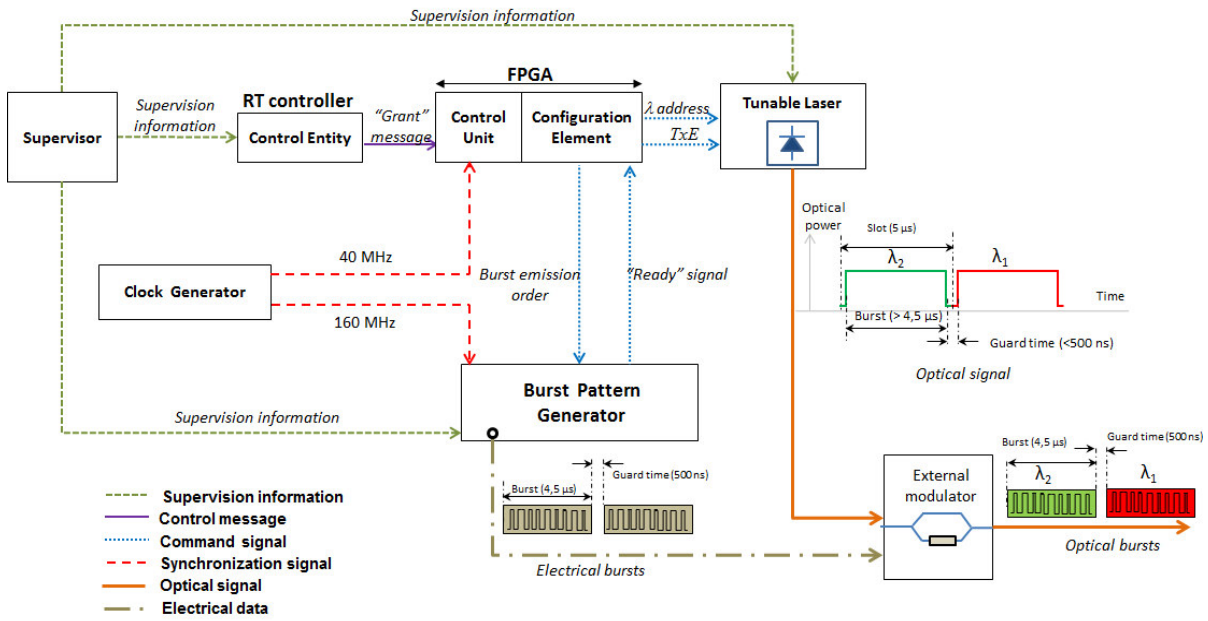


Figure 66- Colored burst generation process

VI.2.1. The supervision function

The supervision function enables the user to monitor the test-bed via the graphical configuration interface provided by the supervisor computer (see Figure 67). As this test-bed emulates the static control plane, the supervisor computer allows the user to configure grant. To do this, the user selects two files containing the emission pattern that the source should follow alternatively: the source uses the first pattern during a given control cycle and then, during the next control cycle, it changes the configuration and uses the second pattern and so

on. The pattern contains 100 items. Each item configures a slot of the data cycle. It has one of these three possible values:

- 1 means that the slot is attributed to the first destination (i.e. using the first wavelength)
- 2 means that the slot is attributed to the second destination (i.e. using the second wavelength)
- 0 means that the slot is unused.

Apart the grant pattern, the user can choose the control cycle duration in milliseconds. According to this duration value, the supervisor computes the number of corresponding data cycles taking into account a fixed data cycle duration equal to 500µs (100 slots of 5 µs each).

The user also determinates the indexes of the wavelengths attributed to the two destinations and the index of the extra wavelength attributed to emit a stuffing burst in the case of unused slot. The purpose of the stuffing burst is to guarantee a continuous signal at the emission side (in order to avoid distorsions at the emission due to the limited low cut-off frequency of the RF components); however these stuffing bursts does not propagate in the network.

Moreover, the user can monitor some internal parameters that vary according to the used devices. They concern the voltage of the signal emitted from the FPGA to the fast tunable laser and also the delay that the FPGA should wait between the reception of the “ready” signal from the BPG and the beginning of the bursts transmission. The “ready” signal informs the FPGA that the BPG is ready to emit electrical bursts.

Finally, the user interface enables the managing of the experiment progress via three buttons “START”, “STOP” and “EXIT”.

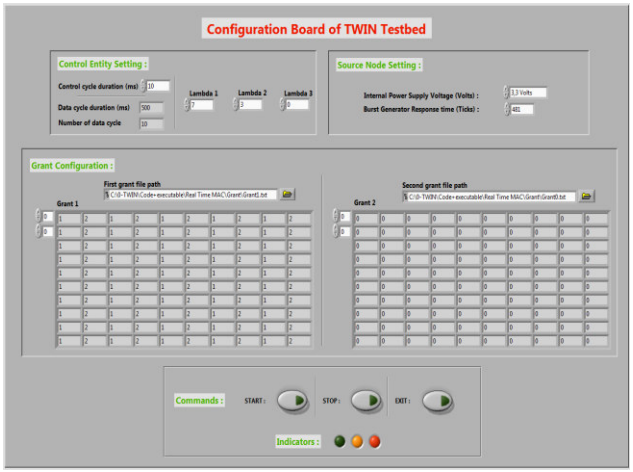


Figure 67- The user interface of the TWIN test-bed configuration board

VI.2.2. The control function

The control function is mainly performed by the control entity module and developed using the LabVIEW real time software. The main role of this module is to ensure the communication with the FPGA that represents the intelligent part of the source node. As depicted in Figure 68, the control entity gets the grant patterns and other setting information from the supervisor in the beginning of the experiment. After memorizing the grant patterns, it establishes the connection with the FPGA module. Every control cycle, the control entity transmits to the FPGA one grant message containing the new burst transmission pattern. The duration of the control cycle is determined by the user via the supervisor computer.

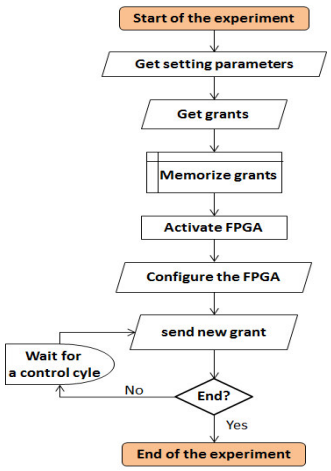


Figure 68- Diagram of the control entity algorithm

VI.2.3. The burst transmission function

The FPGA configures the I/O adapter module to generate, at the right moment, the right signal to the devices that perform the burst transmission, namely, the Burst Pattern Generator (BPG) and the tunable laser

As a preliminary phase, the FPGA activates the Input/Output ports and sets the internal power voltage of the emitted signal then it waits a "ready" signal from the BPG informing that it is well switched on. After receiving the "ready" signal, the FPGA waits for a specific time and then it begins the second phase. The waiting time is set experimentally and manually entered by the user via the supervisor's configuration board. It ensures the time synchronization between the BPG and the FPGA. During this preliminary phase, the FPGA uses its internal clock of 100 MHz.

Once the Input/Output ports are well activated and the BPG is ready to send electrical bursts, the FPGA begins the operating phase. The FPGA manages the electrical burst emitted by the BPG by sending the command signals. Thus, frequency synchronization should be established between these two devices. For this purpose, the FPGA and the BPG are fed by external clocks having a common clock reference provided by the Agilent 81110A. The Agilent 81110A supplies the FPGA with a clock of 40 MHz and the BPG with a clock of 160 MHz as depicted in Figure 66. The BPG module (MU181020A) generates then its proper clock by multiplying the external input clock's rate (160 MHz) by 64. Hence, it obtains a proper clock of 10240 MHz enabling to transmit bursts at 10.24 Gbps.

When the FPGA receives a burst transmission pattern from the control entity, it stores it in an internal memory in order to be repeatedly used in each data cycle. Each item of the pattern is used to configure the BPG and the laser for a slot of 5µs corresponding to 200 occurrences (or ticks) of the 40 MHz clock. Signals generated by the couple FPGA/adapter during these 200 ticks are depicted in Figure 69.

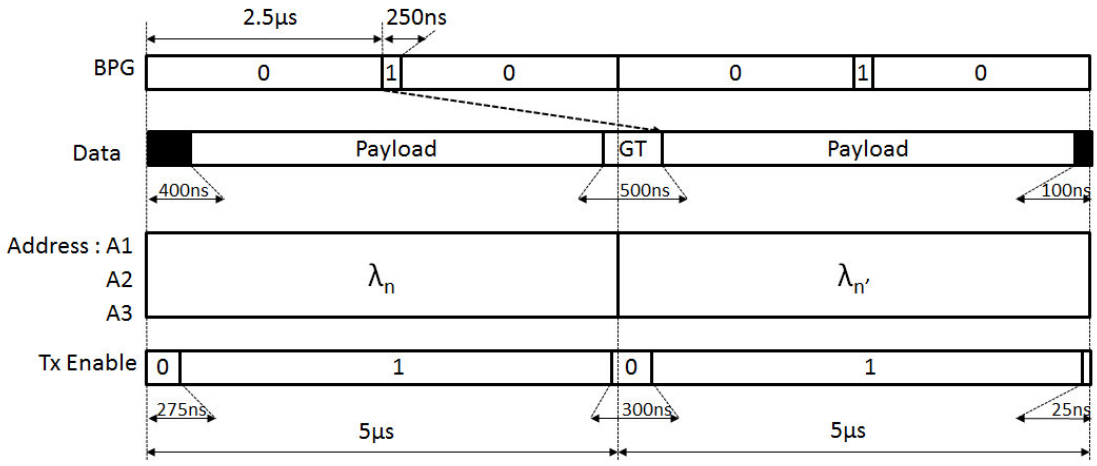


Figure 69- Temporal diagram of the generated signals for burst transmission

The BPG receives a command signal from the FPGA to send an electrical burst. The responsiveness of the BPG to take into account this order and begin the burst emission is equal to 2.9 µs corresponding to 116 ticks. Therefore, the FPGA sends this signal in advance, in the previous slot as shown in Figure 69, in order to ensure the correct timing.

Concerning the tunable laser, the wavelength switching is done under the command of the FPGA in the first tick of each slot. Let's recall here that the command of the wavelength generated by the tunable laser can be handled by an external device via 8 bits address. As the needed wavelength address range is not so large, we design the FPGA to handle only the three least significant bits (A1, A2, and A3). The other bits are kept intact (equal to zero). An index is attributed to each channel as depicted in Table 9. The table also shows the values of the currents required to get the desired wavelengths.

Coding bits			**Channel index**	**Theoretical Wavelength (nm)**	**Currents value (Temperature=29.7°C)**				
A1	**A2**	**A3**			I_{left}**(mA)**	I_{right} **(mA)**	I_{Phase} **(mA)**	I_{SOA} **(mA)**	I_{Gain} **(mA)**
0	0	0	**17**	1558.98	3.6254	3.5798	0.6804	49.0912	99.2
0	0	1	**19**	1558.17	5.4571	5.4567	1.4564	50.6331	99.2
0	1	0	**21**	1557.36	7.8362	7.837	0.3186	50.2075	99.2
0	1	1	**23**	1556.56	11.0801	11.0194	0.8821	51.577	99.2
1	0	0	**25**	1555.75	0.4757	0.877	1.4878	48.3821	99.2
1	0	1	**27**	1554.94	1.0382	1.6275	0.2482	47.4986	99.2
1	1	0	**29**	1554.13	1.8832	2.7366	0.7181	48.8225	99.2
1	1	1	**31**	1553.33	3.0594	4.2447	1.5853	50.1579	99.2

Table 9- Channels parameters

In our test-bed, we choose the following three wavelengths:

- lambda #23: for the burst intended to the first destination,
- lambda #31: for the burst intended to the second destination,
- lambda #17: for the stuffing burst.

The FPGA configures also the TxE port of the laser as shown in Figure 70. When the TxE is set to "1", the laser emits the optical signal. The light is kept switched on during 4.7 µs (188 ticks). Meanwhile, the BPG should emit the electrical burst which has a duration of 4.5 µs. To ensure a good burst transmission, the FPGA has to achieve a perfect synchronization between the tunable laser and the BGP control signals.

When the tunable laser switches from a wavelength to another according to the new available address, the FPGA sets the TxE to "0" during 300 ns. This avoids emitting light and displaying the disturbed optical signal accompanying the wavelength switching process. The duration of this unstable situation depends on the target wavelengths. In the section describing experiments results, we will focus on the way we have determined this switching time for some couple of wavelengths. Both the wavelength switching time and the time needed to turn on/off the laser are among the factors that are taken into account in the determination of the guard time which is the inter-space between two adjacent bursts.

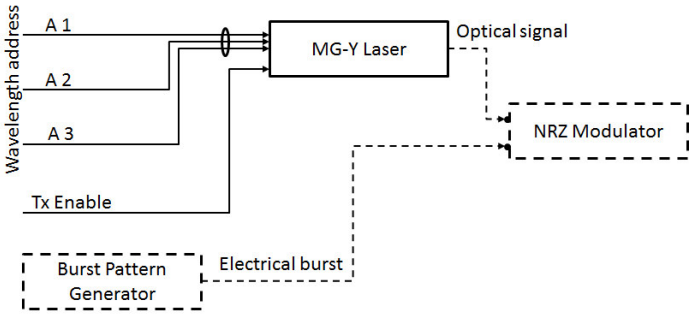


Figure 70- Laser configuration's signals

The electrical burst data coming from the BPG and the optical signal coming from the laser are combined in the external modulator. The electro-optical modulator prints the data distributed by the BGP on the continuous wave provided by the laser.

The transmission burst unit emits stuffing burst when there is no data burst to emit. The use of stuffing data avoids capacitive effect in the modulator driver and thus enables to work

in quasi-continuous regime. To erases the stuffing data in the optical domain, the optical signal exiting the laser passes through an optical filter that rejects the stuffing wavelength.

VI.3. Results

In this section, we present the first results obtained with the demonstrator and some measurement configurations. Figure 71 shows a picture of the demonstrator. It is arranged into two racks. The first rack (on the left) contains the measurement devices, whereas the second rack (on the right) contains the electronic and optical devices composing the test-bed (the indexes of components used in Figure 71 correspond to the indexes used in Figure 61).

The demonstrator is not fully equipped as the source and receiver components are delivered by an external partner and the calendar of the project forecast the delivery of these components mid-2014. However, parts of the demonstrator have been realized and tested. One complete source node has been implemented (with burst generator, tunable laser and modulator). Only the tunable laser is missing in the second one. The core node has been assembled and the control program has been done to manage two source nodes and two destination nodes.

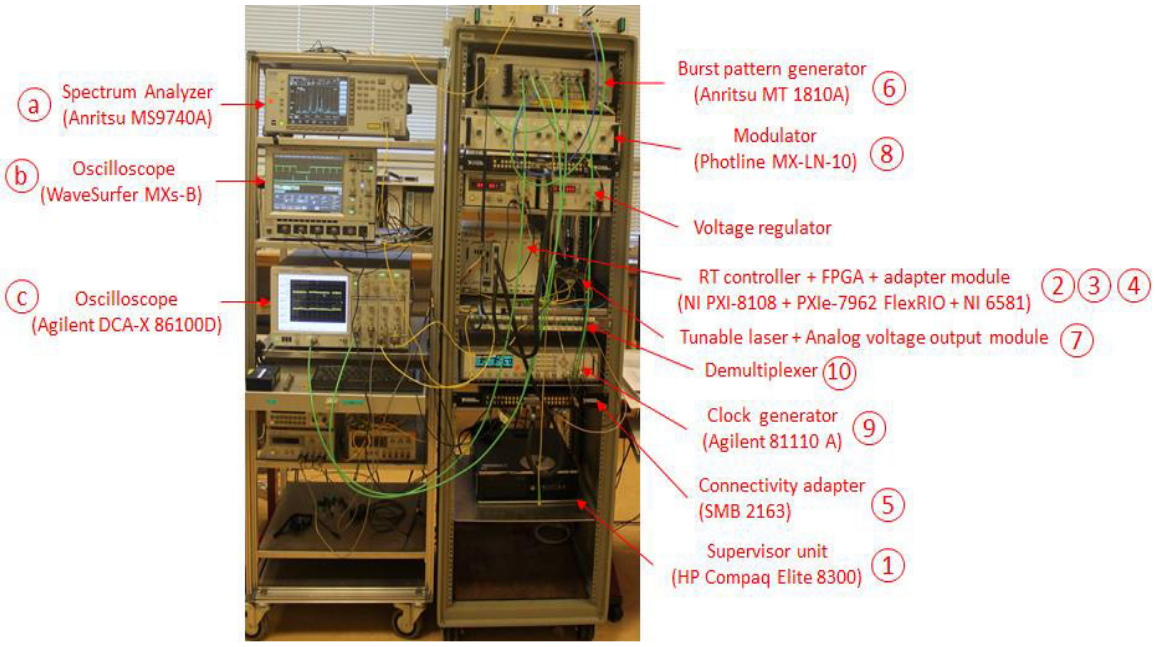


Figure 71- General overview of the test-bed

VI.3.1. Tunable laser generated signal

We use the set-up of Figure 72 to visualize the signal at the output of the laser on the oscilloscope and compute the switching time between two wavelengths. We directly connect the output of the tunable laser to the core node system without passing through the modulator. At the core node, the demultiplexer separates signals according to their wavelengths and directs them individually towards photodectors. The photodector, made from semiconductor materials, converts the optical burst signal into an electrical signal that can be observed via the oscilloscope. In our set-up, we use the oscilloscope LeCroy's WaveSurfer MXs-B [127]. It can capture and perform waveform processing of digital signals of up to 600 MHz. We use this oscilloscope to also observe a copy of some command signals generated by the FPGA in order to verify their voltage and their waveform.

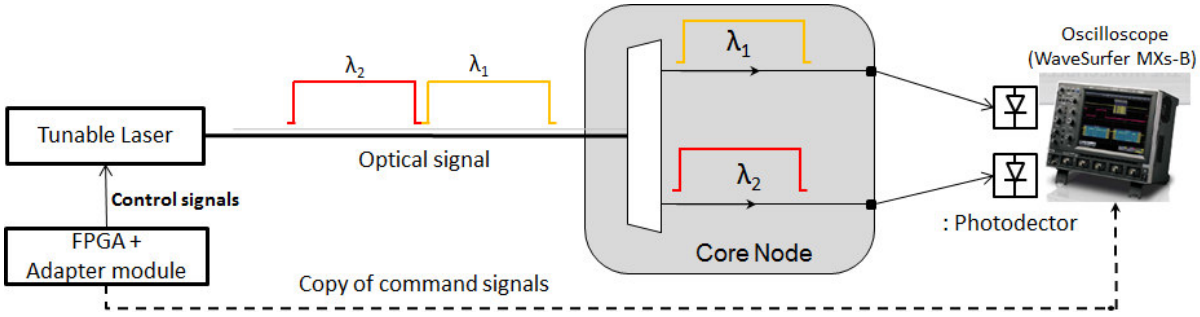


Figure 72- Set-up to evaluate the laser's signal

Before evaluating the optical signal at the output of the tunable laser, we verify the accuracy of the command signals. Therefore, we consider a grant pattern composed of an alternating sequence of "1" and "2". That means that the tunable laser has to switch every 5 µs between Lambda #23 and Lambda #31. The Figure 73 exhibits a green and a blue signals that correspond to A1 and TxE signals respectively. The obtained signals perfectly correspond to the temporal diagram in Figure 69. Indeed, the TxE is set to "0" state 25 ns before the change of the wavelength address. It remains in the "0" binary state for 300 ns while the A1 remains in the "0" state (or in the "1" state) for 5µs.

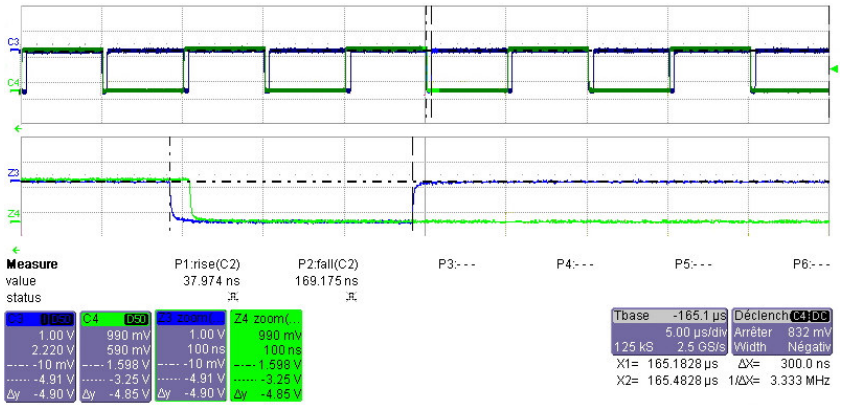
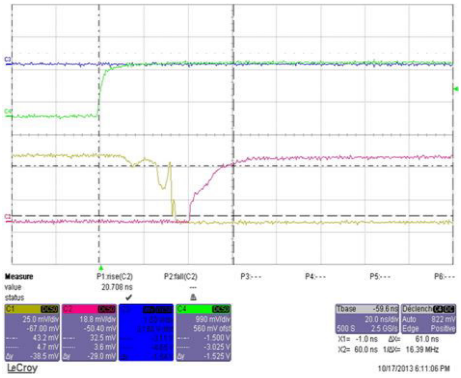
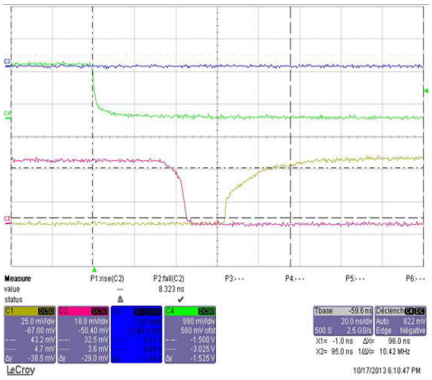


Figure 73- Command signal timing

To assess the wavelength switching time of the laser, we modify, as a first step, the FPGA program in such manner that only the address bits change while the TxE is kept in the "1" binary state. We also try multiple grant patterns. This enables to compute the switching time between the wavelengths. Here, we mean by wavelength switching time, the interval of time between the moment that the FPGA emits the new wavelength address signal and the moment that the new wavelength power reaches the permanent regime. Figure 74, Figure 75 and Figure 76 shows the switching time for all the possible combinations between the three wavelengths (#17, #23 and #31). In these print-screens, the green signal and the blue signal correspond to A1 and TxE respectively (electrical signals), while the red and the yellow signals correspond to lambda #31 and lambda #23 respectively (optical signals).

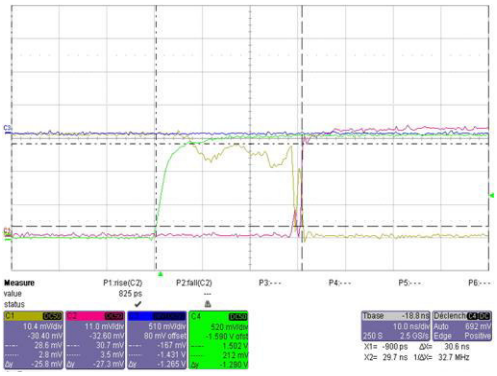


From Lambda #17 to Lambda #23

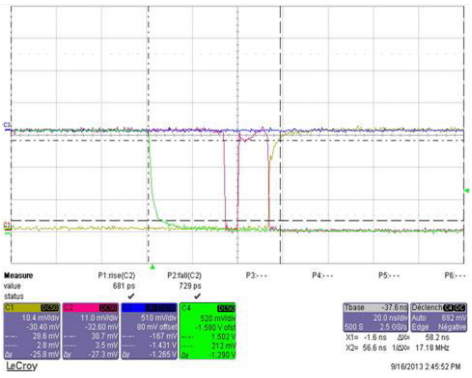


From Lambda #23 to Lambda #17

Figure 74- Switching time between Lambda #17 and Lambda #23

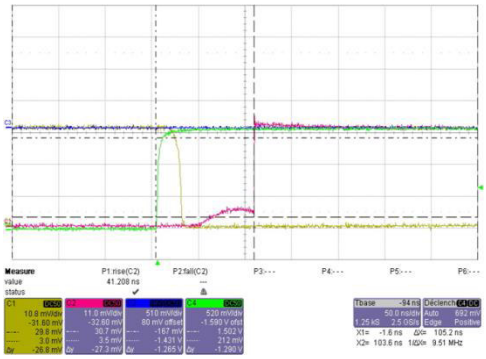


From Lambda #17 to Lambda #31

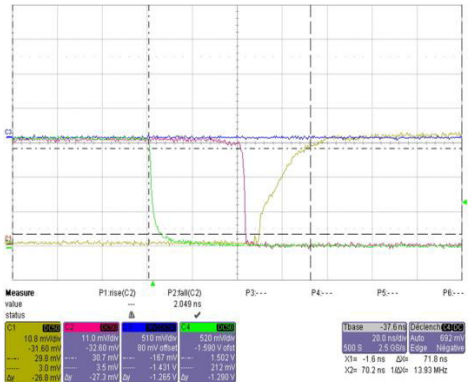


From Lambda #31 to Lambda #17

Figure 75- Switching time between Lambda #31 and Lambda #17



From Lambda #23 to Lambda #31



From Lambda #31 to Lambda #23

Figure 76- Switching time between Lambda #31 and Lambda #23

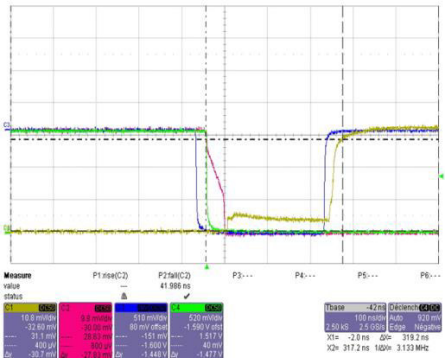
The results depicted in the previous figures and in the summary of the Table 10 show that the wavelength switching times do not exceed 105 ns for the three wavelengths.

Ongoing wavelength index	Target wavelength index	Switching time (ns)
17	23	61
23	17	96
17	31	30
31	17	58
23	31	105
31	23	71

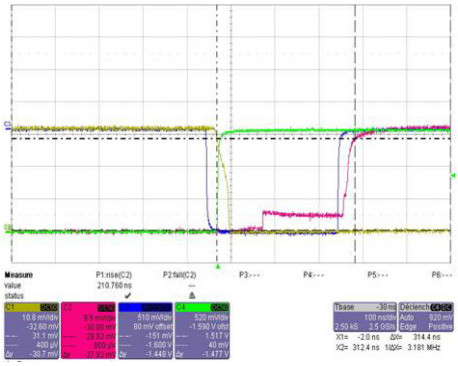
Table 10- Summary of wavelength switching times (ns)

As a second step, we return back to the original version of the FPGA program and we set the TxE to "0" during the wavelength switching. The Figure 77 depicts the aforementioned signals when the laser switches between Lambda #31 and Lambda #23. The signal perturbation accompanying this process is hidden by the TxE that switches off the light emitted by the laser. To do this, the TxE drives the SOA current. Setting the SOA current to

"0" does not completely switch off the laser. The SOA section should be reverse bias but this is not done in the current electrical card. Hence, a slight yellow signal appears. After 300 ns, the TxE is set to 1 and the yellow signal rises again and reaches 90% of the maximum after 50 ns. The time between the falling of the lambda #31 and the rising of lambda #23 is equal to 319 ns. The same process takes 314.4 ns to switch from lambda #23 to lambda #31. In the two cases, the switching time process is still inferior to the guard time (500 ns).



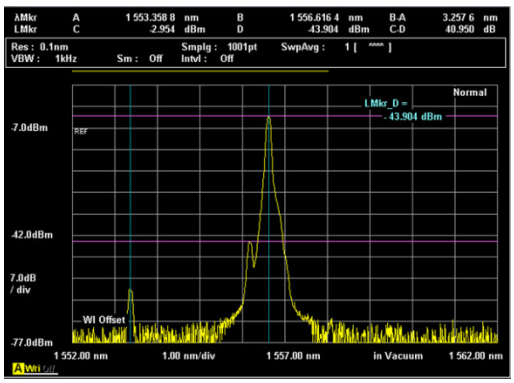
From Lambda #31 to Lambda #23



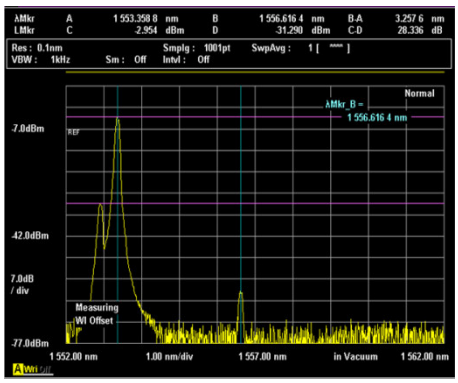
From Lambda #31 to Lambda #23

Figure 77- Switching between Lambda #31 and lambda #23

To assess the quality of signals, we place the Optical Spectrum Analyzer (OSA) in the output of the core node. The OSA is the MS9740A of Anritsu [128]. The obtained results, shown in Figure 78, indicate that Lambda #23 and Lambda #31 in the output of the demultiplexer have a wavelength equal to 1556.61 nm and 1553.35 nm respectively and their power is equal to -7 dBm. The measured wavelengths are close to the theoretical values which are equal to 1556.56 nm for Lambda #23 and 1553.33 nm for Lambda #31.



Lambda #23



Lambda #31

Figure 78- Spectrum analyze of the switching between Lambda #23 and Lambda #31

VI.3.2. Modulated signal

In this section, we assess the optical signal generated by the NRZ modulator. Using the set-up depicted in Figure 79, we measure the eye opening and detect waveform transitions that cross through the eye. In this set-up, the modulator is directly related to a powerful oscilloscope (86100D DCA-X). This high-speed sampling oscilloscope combines high analog bandwidth, low jitter, and low noise performance to accurately characterize optical and electrical signals from 50 Mbps to over 80 Gbps. In our experiments, we use it to display the eye diagram that is generated by applying a synchronized superposition of all possible realizations of the bit stream signal. As we use a NRZ modulation format, the transition between bits equal to "1" and others equal to "0" corresponds to rising or falling of the signal. The open spaces, seen between these transitions, present the eye. The degree of eye opening indicates the signal quality. For instance, in the case of signal waveform distortion due to inter-symbol interference or noise, a closure of the eye pattern appears.

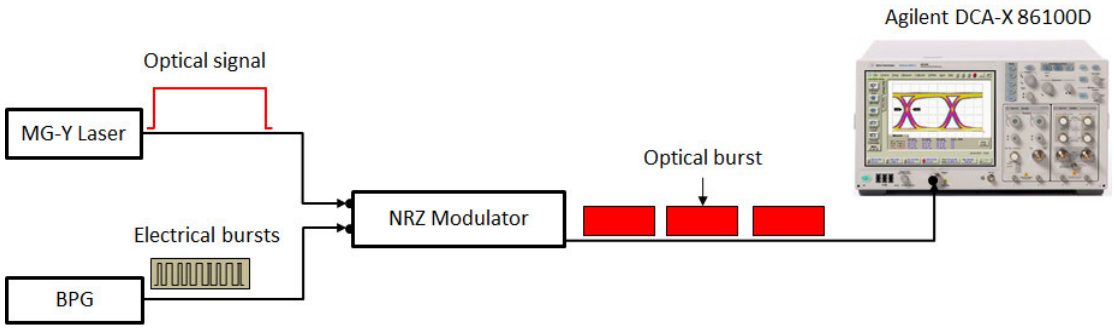


Figure 79- Set-up to evaluate the modulator

The obtained eye diagram of an output signal having a peak to peak voltage equals to 0.4 is shown in Figure 80. The displayed "open eye" indicates low bit error rates and minimum signal distortion. The maximum rise time is equal to 36.9 ps, while the maximum fall time is equal to 35.1 ps. The peak-to-peak jitter, corresponding to the difference between the extreme right point and the left point of the eye, exhibits a maximum value of 12.89 ps.

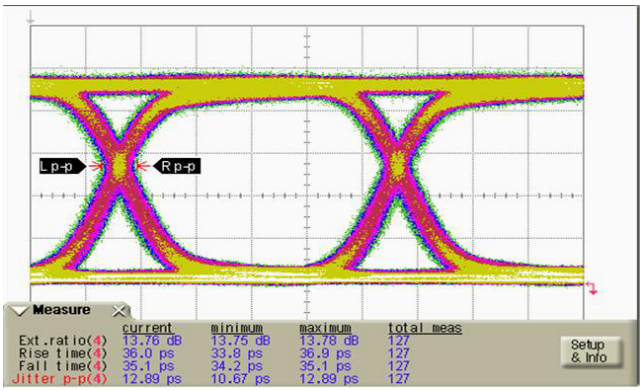


Figure 80- Eye diagram

VI.3.3. Burst signal

We use the measurement set-up of Figure 81 to visualize the optical bursts in the output of the modulator and assess the accuracy of its emission time and the quality of its optical signal. Therefore, when the burst exits the emission system, it passes through a coupler that splits the incoming optical signal into two parts. The first part having 10% of the total power is transmitted to the spectrum analyzer that measures the power distribution of optical wavelengths while the second part, placed on the output 2 of the coupler, is transmitted to the core node. After being filtered by the demultiplexer, bursts are received by the photodectors connected to the oscilloscope. As in the previous set-up, a copy of some command signals generated by the FPGA is also displayed on the oscilloscope traces.

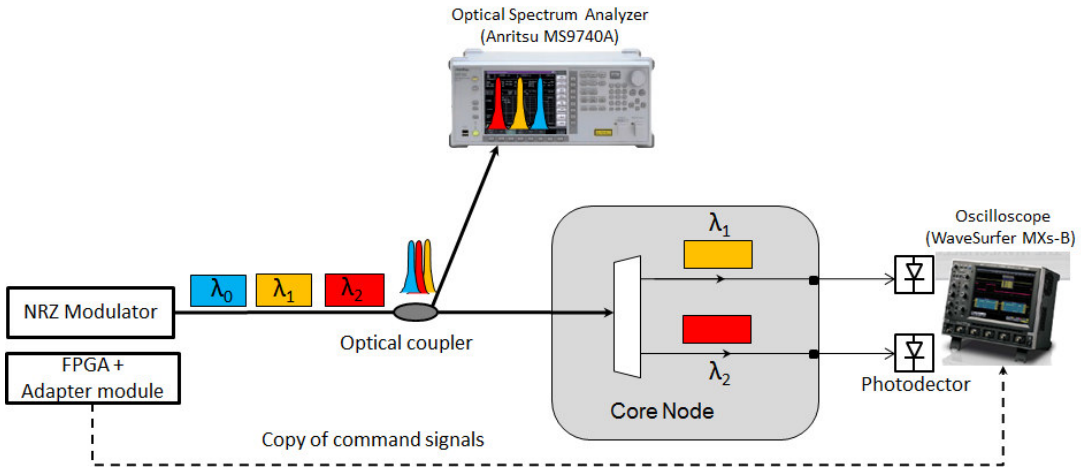


Figure 81- Set-up to evaluate the obtained bursts

The upper diagram of the Figure 82 displays two command signals configuring the laser (the address signal (A1) in green and the TxE signal in blue) and the corresponding bursts. According to the used burst emission pattern, the number of stuffing bursts to be emitted by the source is the half of the number of bursts intended to each destination. At the bottom of Figure 82, we see a zoom of the inter-burst. We verify that the burst duration is equal to 4.5µs and the guard time between two successive bursts is equal to 500 ns.

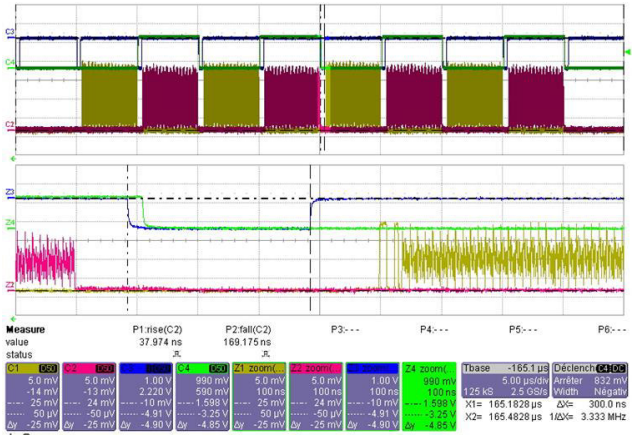


Figure 82- Bursts and command signals

Figure 83 shows a print-screen for the OSA and the obtained results, concerning the three optical channels of the system (#17, #23 and #31), are collected in Table 11.

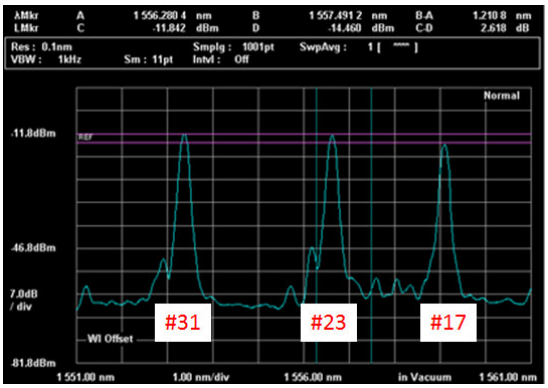


Figure 83- Wavelengths spectrum

Results show that the measured wavelengths are close to the theoretical value. The difference is lower than 0.1 nm due to uncertainties in the setting currents and also the switching process that induces wavelength drifts. Channels # 23 and #31 have the same power, which is equal to -11.8 dBm. However, the channel #17 power is equal to -14.4 dBm.

This difference of almost 3dB is due to the data cycle configuration considered in this experiment. In fact, the number of stuffing bursts is the half of the number of bursts intended to each destination. Hence, the power of the data bursts is barely the double of the power of the stuffing bursts.

Channel number	Theoretical wavelength (nm)	Measured wavelength (nm)	Difference (nm)	Power (dBm)
17	1558.98	1559.08	0.1	-14.4
23	1556.56	1556.63	0.07	-11.8
31	1553.33	1553.35	0.02	-11.8

Table 11- Channels characteristics

VI.4. Conclusion

In this chapter, we have described the test-bed that we designed as a PoC of the TWIN paradigm. It demonstrates the feasibility of some critical parts of TWIN technology such as the fast wavelength switching and burst emission process and it also permits to validate some hypothesis concerning the guard time and the burst duration. The implementation of the intelligent features of this test-bed is mainly based on National Instruments devices through the use of a powerful real time target and sophisticated FPGA card.

The main challenges of this test-bed are the overcome of the high temporal constraints to reach a data transmission bit rate of 10 Gbps and the complexity of establishing a perfect synchronization between the different components. Despite these constraints, the static control plane and the burst mode transmitter were successfully realized and tested and the obtained results show the accuracy of the generated signals and their compliance with the theoretical study.

Compared with the test-bed done in Shanghai Jiao Tang University and described in [108], our test-bed includes a sophisticated burst emission unit operating at 10 Gbps instead of 1.25 Gbps. Furthermore, a separate control entity is designed to execute a static control plane that manages the burst emission pattern. This test-bed provides also supervision features enabling the user to configure and control the entire system. The conception of the algorithms and the flexibility of the embedded devices offer the possibility to upgrade the system by

adding other sources and destinations or modifying the control plane mechanisms and parameters.

This test-bed continues to evolve in the frame of SASER project; especially it will be fully equipped in order to build the two sources and two destination nodes. As future work, we aim to implement a dynamic control plane to allocate resources during a control cycle of several milliseconds. According to this control plane, sources send to the control entity request messages containing randomly generated resource needs. Taking into account these requests, the control entity executes a dynamic resource allocation algorithm, implemented on the real-time target, to compute grants for the next control cycle.

The other important component planned to this test-bed is the burst mode receiver that will enable to decode bursts and assess the Bit Error Rate (BER). As a preliminary model of this component, we intend to add stuffing data at the reception side to fill in the gap between bursts in such manner that the receiver will work in quasi continuous mode. The receiver could benefit from the grants generated by the control entity to know the time slots when it will not receive data. This information would help the receiver to estimate the right time to add stuffing data. The quasi continuous mode receiver could conserve the clock rhythm from one burst to another. However, having the right frequency is not enough to ensure an efficient burst reception; the phase shall be also retrieved. This can be done by adding a preamble in the beginning of the burst as synchronization symbols. Hence, the time response in phase is among the parameters that should be evaluated using different guard time intervals.

Chapitre VII. Conclusion and perspectives

Optical burst switching was defined more than fifteen years ago and a significant amount of research has been done on this subject since then. Given the large corpus of literature on this subject, we chose to study in depth one of the proposed solution instead of designing a new one. After a detailed description of the existing alternatives in chapter 3, we have selected the TWIN solution as it provides lossless burst switching with simple and all-optical intermediate nodes, within a mesh topology. The simplicity of TWIN nodes comes at the expense of a complex control plane to avoid burst contention.

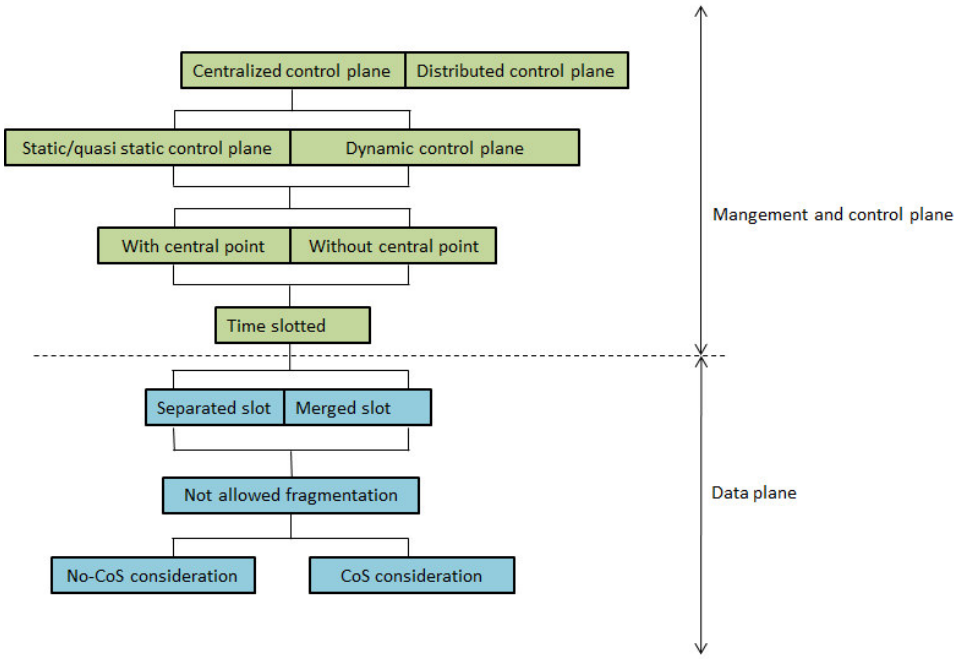


Figure 84- Summary of TWIN studied features

Through this study, we proposed several mechanisms to perform the management/control plane and the data plane of TWIN as depicted in Figure 84. Mechanisms concerning the management/control plane mainly depend on the entity that computes the allocation of resources (centralized or distributed), the reactivity of the control plane (static or dynamic) and the topology of the network (with central point or without central point). However,

mechanisms concerning the data plane are mainly related to the way the time slots are used (separated or merged slots), the frame fragmentation and the CoS consideration.

Taking into account this structure, we first compare in chapter 4 the centralized and the distributed control plane in terms of end-to-end delay, jitter, queue length and wavelength utilization. Simulation results prove that in an aligned case the centralized scheme outperforms the distributed scheme by almost 15%. Based on these results, we compared the performance of different centralized control planes. The centralized control planes are either dynamic or static. The dynamic schemes are based on a heuristic approach to perform the resource allocation (i.e. scheduler computation) which changes according to the traffic variation observed during a short period (a "control cycle" of several milliseconds duration). On the other hand, the static approach is based on an optimized resource allocation based on a traffic matrix and the resource allocation is kept unchanged during a significant period (from several seconds to several minutes). The results obtained by considering synthetic traffic profiles during the simulations show that the static scheme allows a bandwidth utilization of more than 80% and it performs better than the dynamic schemes. In order to confirm these preliminary results and verify the robustness of the static scheme, we have used real traffic traces in chapter 5 to drive our simulations. To the best of our knowledge, this is one of the few contributions on this topic that uses real traffic traces instead of synthetic traffic models. Next, we have proposed to apply TWIN to a new architecture, MEET, which is intended for a metro-backhaul network. This architecture does not only flatten the current "hub-and-spoke" architecture but also improves resource utilization since it implicitly ensures slot alignment thanks to its central-point based architecture. Results show that despite of the high variation of the real traffic, the static scheme still performs well. We also prove that the centralized control plane of MEET could be coupled with a QoS-aware burst assembly mechanism in order to differentiate traffic and in particular to favor delay-sensitive traffic.

Although they present real technical advantages, sub-wavelength switching solutions often suffer from the lack of necessary infrastructure technologies. Therefore, they are considered as immature for the time being and the near future. Therefore, designing a PoC for proposed sub-wavelength switching solutions is important. In the test-bed described in chapter 6, we proved that developing TWIN nodes with static control plane is actually

feasible with the following selected parameters: burst size equal to 5µs, guard time equal to 0.5µs and 10 Gbps rate links. Despite the lack of available components, we have succeeded in ensuring synchronization between the different parts of the test-bed and in obtaining the correct time accuracy with optical signals of good quality.

This work was partly done in the frame of the CELTIC-Plus project SASER-SaveNet and is still evolving. As next steps, we aim to carry out comparative study with other control plane proposed by our partners in the aforementioned project. It would be particularly interesting to evaluate their performance using different traffic traces and scenarios.

As future work, we also intend to further investigate the potential of sub-wavelength switching solution, specifically TWIN, to face the new optical transport technologies that are gaining great momentum. Flexgrid networks [129] are among these trendy technologies that are attracting huge interest due to their higher spectrum efficiency and flexibility. Moreover, we will explore the capability of Software Defined Network (SDN) [130] to implement a flexible control plane for sub-wavelength solutions. SDN should be able to provide some interesting functionality such as service setup and teardown, service parameter modification and service events and alarming to a TWIN infrastructure.

References

[1] ITU-T, «Recommandation G.803 : Architectures of transport networks based on the Synchronous Digita Hierarchy (SDH)», 2000.

[2] M. J. O'Mahony, D. Simeonidou, D. K. Hunter et A. Tzanakaki, «The application of optical packet switching in future communication networks», *Communications Magazine, IEEE,* vol. 39, n° 3, pp. 128-135, 2001.

[3] C. Qiao et M. Yoo, «Optical Burst Switching (OBS)-a new paradigm for an optical Internet», *Journal of high speed networks,* vol. 8, n° 1, p. 69, 1999.

[4] ITU-T, «Terms and definitions for Sub-Lambda Photonically Switched Networks», Geneva, 2012.

[5] I. Widjaja, I. Saniee, R. Giles et D. Mitra, «Light core and intelligent edge for a flexible, thin-layered, and cost-effective optical transport network», *Communications Magazine, IEEE,* vol. 41, n° 5, pp. 30-36, 2003.

[6] E. Bonetto, A. Triki, E. Le Rouzic, B. Arzur et P. Gavignet, «Circuit switching and time-domain optical sub-wavelength switching technologies: Evaluations on the power consumption», *20th International Conference on Software, Telecommunications and Computer Networks (SoftCOM),* pp. 1-5, 2012.

[7] A. Triki, P. Gavignet, B. Arzur, E. Le Rouzic et A. Gravey, «Efficient control plane for passive optical burst switching network», *International Conference on Information Networking (ICOIN),* pp. 535-540, 2013.

[8] A. Triki, P. Gavignet, B. Arzur, E. Le Rouzic et A. Gravey, «Bandwidth allocation schemes for a lossless Optical Burst Switching.», *17th International Conference on Optical Network Design and Modeling (ONDM),* pp. 205-210, 2013.

[9] R. Aparicio-Pardo, A. Triki, E. Le Rouzic, B. Arzur, E. Pincemin et F. Guillemin, «Alternate architectures for an all-optical core network based on new subwavelength switching paradigms», *15th International Conference on Transparent Optical Networks (ICTON),* pp. 1-4, 2013.

[10] A. Triki, P. Gavignet, B. Arzur, E. Le Rouzic et A. Gravey, «Is It worth adapting sub-wavelength switching control plane to traffic variations?», *18th International Conference on Optical Network Design and Modeling (ONDM),* 2014.

[11] D. Hardy, G. Malléus et J.-N. Méreur, Réseaux: Internet, téléphonie, multimédia. Convergences et complémentarités, De Boeck Supérieur, 2002.

[12] ITU-T, «Recommandation G.984 : Gigabit-capable Passive Optical Networks (GPON)», 2008.

[13] M. Feknous, B. Le Guyader et A. Gravey, «Revisiting access and aggregation network architecture», *The International conference on communication systems and computational intelligence,* 2014.

[14] ITU-T, «Recommendation G.987 : 10 Gigabit capable Passive Optical Networks (XGPON)», 2010.

[15] ITU-T, «Recommendation G.983.1: Broadband Optical access system based on Passive Optical Networks (GPON)», 2005.

[16] IEEE Std 802.3av, *Specific requirements-- Part 3: CSMA/CD Access Method and Physical Layer Specifications Amendment 1: Physical Layer Specifications and Management Parameters for 10 Gb/s Passive Optical Networks,* 2009.

[17] O. Haran et A. Sheffer, «The importance of dynamic bandwidth allocation in GPON networks», *PMC-Sierra Inc., White Paper,* n° 1, 2008.

[18] H.-C. Leligou, C. Linardakis, K. Kanonakis, J. D. Angelopoulos et T. Orphanoudakis, «Efficient medium arbitration of FSAN-compliant GPONs», *international journal of communication systems,* vol. 19, n° 5, pp. 603-617, 2006.

[19] B. Skubic, J. Chen, J. Ahmed, L. Wosinska et B. Mukherjee, «A comparison of dynamic bandwidth allocation for EPON, GPON, and next-generation TDM PON», *Communications Magazine, IEEE,* vol. 47, n° 3, pp. 40-48, 2009.

[20] 3GPP TR 45.912, «Feasibility study for evolved GSM/EDGE Radio Access Network (GERAN)», 2009.

[21] 3GPP TS 25.401, «UTRAN Overall Description», 2011.

[22] 3GPP TS 36.300, «E-UTRA and E-UTRAN; overall description», 2012.

[23] Cisco Systems, «CISCO Visual and Index Forecast and Methodology, 2012-2017», 2013.

[24] H. Al-Raweshidy et S. Komaki, Radio over fiber technologies for mobile communications networks, Artech House, 2002.

[25] ITU-T, «Recommandation I.361 : B-ISDN ATM layer specification», 1999.

[26] P. Gravey, A. Gravey, M. Morvan, L. Sadeghioon et B. Uscumlic, «Status of time-slotted optical packet-switching and its application to flexible metropolitan networks», *2ème colloque réseau large bande et internet rapide,* 2011.

[27] IEEE Std 802.3ba, *Amendment 4: Media Access control parameters, physical layers and management parameters for 40 Gb/s and 100 Gb/s operation,* 2010.

[28] P. Walker, «Optical Transport Network (OTN) tutorial», 2008.

[29] E. Hernandez-Valencia, M. Scholten et Z. Zhu, «The generic framing procedure (GFP): An overview», *Communications Magazine, IEEE,* vol. 40, n° 5, pp. 63-71, 2002.

[30] V. Eramo, M. Listanti, R. Sabella et F. Testa, «Definition and performance evaluation of a low-cost/high-capacity scalable integrated OTN/WDM switch», *Journal of Optical Communications and Networking,* vol. 4, n° 12, pp. 1033-1045, 2012.

[31] K. H. Liu, IP over WDM, vol. 389, John Wiley and Sons, 2002.

[32] N. Ghani, S. Dixit et T.-S. Wang, «On IP-over-WDM integration», *Communications Magazine, IEEE,* vol. 38, pp. 72-84, 2000.

[33] M. Kodialam et T. Lakshman, «Integrated dynamic IP and wavelength routing in IP over WDM networks», *20th Annual IEEE International Conference on Computer Communications (INFOCOM),* vol. 1, pp. 358-366, 2001.

[34] D. Awduche et Y. Rekhter, «Multiprotocol lambda switching: combining MPLS traffic engineering control with optical crossconnects», *Communications Magazine, IEEE,* vol. 39, pp. 111-116, 2001.

[35] M. Murata et K.-i. Kitayama, «A perspective on photonic multiprotocol label switching», *Network, IEEE,* vol. 15, pp. 56-63, 2001.

[36] G. Roberts et G. Amato, «The G.709 Optical Transport Network: An overview», 2006.

[37] IEEE Std 802.3, *Information technology - telecommunications and information exchange between systems - local and metropolitan area networks - specific requirements. Part 3: Carrier Sense Multiple Access with Collision Detection (CSMA/CD) access method and physical layer,* 1998.

[38] IEEE Std 802.3z, *Media Access Control (MAC) Parameters, Physical Layer, Repeater and Management Parameters for 1000 Mbps Operation,* P. N. IEEE Press, Éd., 1998.

[39] IEEE Std 802.3x, *Full Duplex Operation,* P. N. IEEE Press, Éd., 1997.

[40] IEEE Std 802.1Q, *Local and metropolitan area networks-Media Access Control (MAC) Bridges and Virtual Bridged Local Area Networks,* 2011.

[41] IEEE Std 802.1ad, *Local and Metropolitan Area Networks--Virtual Bridged Local Area Networks--Amendment 4: Provider Bridges,* 2005.

[42] IEEE Std 802.3ah, *IEEE 802.1ah, "Virtual Bridged Local Area Networks, Amendment 6: Provider Backbone Bridges,* 2008.

[43] K. a. R. Y. Kompella, «Virtual private LAN service (VPLS) using BGP for auto-discovery and signaling», 2007.

[44] M. a. K. V. Lasserre, «Virtual private LAN service (VPLS) using label distribution protocol (LDP) signaling», 2007.

[45] J. V. Mocerino, «Carrier class Ethernet service delivery migrating SONET to IP and triple play offerings», *National Fiber Optic Engineers Conference (NFOEC),* p. JThB97, 2006.

[46] L. Martini, E. Rosen, N. El-Aawar et G. Heron, «Encapsulation methods for transport of Ethernet over MPLS networks», 2006.

[47] J. Postel, «Internet protocol -DARPA Internet Program Protocol Specification RFC 791», *USC/Information Sciences Institute, USC/Information Sciences Institute,* 1981.

[48] K. Nichols, S. Blake, F. Baker et D. Black, «Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers», 1998.

[49] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang et W. Weiss, «An architecture for differentiated services», 1998.

[50] J. Postel, «Transmission control protocol -DARPA Internet Program Protocol Specification RFC 793», *USC/Information Sciences Institute, USC/Information Sciences Institute,* 1981.

[51] J. Postel, «User datagram protocol -DARPA Internet Program Protocol Specification RFC 768», *USC/Information Sciences Institute, USC/Information Sciences Institute,* 1980.

[52] J. Yan et E. Chen, «Autonomous-System-Wide Unique BGP Identifier for BGP-4», 2011.

[53] J. Davies, Understanding IPv6, Microsoft Press, 2010.

[54] R. M. Hinden et S. E. Deering, «IP version 6 addressing architecture», 2006.

[55] E. Rosen, A. Viswanathan et R. Callon, «Multiprotocol label switching architecture», 2001.

[56] D. Beller et R. Sperber, «MPLS-TP-The New Technology for Packet Transport Networks», *DFN-Forum Kommunikationstechnologien,* pp. 81-92, 2009.

[57] E. Mannie, «Generalized multi-protocol label switching (GMPLS) architecture», 2004.

[58] A. Farrel et I. Bryskin, GMPLS: architecture and applications, Morgan Kaufmann, 2006.

[59] Bell Labs, «Metro network traffic growth: an architecture impact study», 2013.

[60] S. Blouza, J. Karaki, N. Brochier, E. Le Rouzic, E. Pincemin et B. Cousin, «Multi-band OFDM for optical networking», *EUROCON-International Conference on Computer as a Tool (EUROCON), 2011 IEEE,* pp. 1-4, 2011.

[61] J. M. Simmons, «Network design in realistic" all-optical" backbone networks», *Communications Magazine, IEEE,* vol. 44, n° 11, pp. 88-94, 2006.

[62] A. Sen, S. Murthy et S. Bandyopadhyay, «On sparse placement of regenerator nodes in translucent optical network», *Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008. IEEE,* pp. 1-6, 2008.

[63] O. Pedrola, D. Careglio, M. Klinkowski et J. S. Pareta, «Modelling and performance

evaluation of a translucent OBS network architecture», *Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE,* pp. 1-6, 2010.

[64] Y. Xiong, M. Vandenhoute et H. Cankaya, «Control architecture in optical burst-switched WDM networks», *Selected Areas in Communications, IEEE Journal on,* vol. 18, n° 10, pp. 1838-1851, 2000.

[65] X. Yu, Y. Chen et C. Qiao, «Study of traffic statistics of assembled burst traffic in optical burst-switched networks», *ITCom 2002: The Convergence of Information Technologies and Communications,* pp. 149-159, 2002.

[66] J. Wei et R. McFarland, «Just-in-time signaling for WDM optical burst switching networks», *Journal of Lightwave Technology,* vol. 18, n° 12, pp. 2019-2037, 2000.

[67] M. Yoo et C. Qiao, «Just-Enough-Time (JET): A high speed protocol for bursty traffic in optical networks», *IEEE/LEOS Summer Topic Meeting,* pp. 26-27, 1997.

[68] T. Legrand, H. Nakajima, P. Gavignet, B. Charbonnier et B. Cousin, «Etude numérique de la résolution spectro-temporelle de contention de burst et réalisation d'un noeud OBS», *JNOG,* 2008.

[69] T. Legrand, H. Nakajima, P. Gavignet et B. Cousin, «Comparaison de l'OBS conventionnel et de l'OBS à label», *submitted to JNOG,* 2008.

[70] T. Legrand, H. Nakajima, P. Gavignet et B. Cousin, «Labelled OBS test bed for contention resolution study», *5th International Conference on Broadband Communications, Networks and Systems (BROADNETS),* pp. 82-87, 2008.

[71] T. Legrand, B. Cousin et N. Brochier, «Performance evaluation of the labelled OBS architecture», *7th International Conference on Wireless And Optical Communications Networks (WOCN),* pp. 1-6, 2010.

[72] D. Chiaroni, «Optical packet add/drop multiplexers for packet ring networks», *34th European Conference on Optical Communication (ECOC),* pp. 1-4, 2008.

[73] D. Chiaroni, G. Buforn, C. Simonneau, S. Etienne et J. Antona, «Optical packet add/drop systems», *Optical Fiber Communication Conference (OFC),* pp. 1-3, 2010.

[74] C. Cadere, N. Izri, D. Barth, J. Fourneau, D. Marinca et S. Vial, «Virtual circuit allocation with QoS guarantees in the ECOFRAME optical ring», *14th Conference on Optical Network Design and Modeling (ONDM),* pp. 1-6, 2010.

[75] D. Chiaroni, C. Simonneau, M. Salsi, G. Buforn, S. Etienne, H. Mardoyan, J. Simsarian et J. Antona, «Optical packet ring network offering bit rate and modulation formats transparency», *Optical Fiber Communication Conference (OFC),* 2010.

[76] L. Sadeghioon, A. Gravey et P. Gravey, «A label based MAC for OPS multi-rings», *15th International Conference on Optical Network Design and Modeling (ONDM),* pp. 1-6, 2011.

[77] D. Chiaroni, R. Urata, J. Gripp, J. Simsarian, G. Austin, S. Etienne, T. Segawa, Y.

Pointurier, C. Simonneau, Y. Suzaki et others, «Demonstration of the interconnection of two optical packet rings with a hybrid optoelectronic packet router», *36th European Conference and Exhibition on Optical Communication (ECOC),* pp. 1-3, 2010.

[78] T. Bonald, S. Oueslati, J. Roberts et C. Roger, «SWING: Traffic capacity of a simple WDM ring network», *21st International Teletraffic Congress (ITC),* pp. 1-8, 2009.

[79] T. Eido, F. Pekergin et T. Atmaca, «Performance analysis of an enhanced distributed access mechanism in a novel multiservice OPS architecture», *Next Generation Internet Networks (NGI),* pp. 1-7, 2009.

[80] B. Uscumlic, A. Gravey, I. Cerutti, P. Gravey et M. Morvan, «Stable optimal design of an optical packet ring with tunable transmitters and fixed receivers», *Optical Network Design and Modeling (ONDM), 2013 17th International Conference on,* pp. 82-87, 2013.

[81] S. Cao, N. Deng, T. Ma, J. Qi, X. Shi, J. He et J. Zhou, «An optical burst ring network featuring sub-wavelength- and wavelength-granularity grooming», *Photonics Global Conference (PGC),* pp. 1-3, dec. 2010.

[82] S. Cao, N. Deng, X. Shi, T. Ma, Q. Xue, S. Xu et N. Srinivasan, «Optical Burst Transport Network», *37th European Conference and Exhibition on Optical Communication (ECOC),* 2011.

[83] N. Deng, S. Cao, T. Ma, X. Shi, X. Luo, S. Shen et Q. Xiong, «A novel optical burst ring network with optical-layer aggregation and flexible bandwidth provisioning», *Optical Fiber Communication Conference (OFC),* 2011.

[84] J. Fernandez-Palacios, L. Perez, J. Rodriguez, J. Dunne et M. Basham, «IP offloading over multi-granular photonic switching technologies», *36th European Conference and Exhibition on Optical Communication (ECOC),* pp. 1-3, 2010.

[85] J. Dunne, T. Farrell et J. Shields, «Optical Packet Switch and Transport: A new metro platform to reduce costs and power by 50\% to 75\% while simultaneously increasing deterministic performance levels», *11th International Conference on Transparent Optical Networks (ICTON) ,* pp. 1-5, 2009.

[86] J. Fernandez-Palacios, N. Gutierrez, G. Carrozzo, G. Bernini, J. Aracil, V. Lopez, G. Zervas, R. Nejabati, D. Simeonidou, M. Basham et others, «Metro architectures enabliNg subwavelengths: rationale and technical challenges», *Future Network and Mobile Summit,* pp. 1-8, 2010.

[87] C. Kallo, M. Basham, J. Dunne, J. Fernandez-Palacios et others, «Cost reduction of 80% in next-generation virtual personal computer service economics using a sub-wavelength metro network», *16th European Conference on Networks and Optical Communications (NOC),* pp. 224-227, 2011.

[88] M. Duser et P. Bayvel, «Analysis of a dynamically wavelength-routed optical burst switched network architecture», *Lightwave Technology, Journal of,* vol. 20, n° 4, pp. 574-585, 2002.

[89] M. Duser et P. Bayvel, «Performance of a dynamically wavelength-routed optical burst switched network», *Photonics Technology Letters, IEEE,* vol. 14, n° 2, pp. 239-241, 2002.

[90] E. Kozlovski, M. Duser, I. De Miguel et P. Bayvel, «Analysis of burst scheduling for dynamic wavelength assignment in optical burst-switched networks», *The 14th Annual Meeting of the IEEE Lasers and Electro-Optics Society (LEOS) ,* vol. 1, pp. 161-162, 2002.

[91] M. Duser, I. De Miguel, P. Bayvel et D. Wischik, «Timescale analysis for wavelength-routed optical burst-switched (WR-OBS) networks», *Optical Fiber Communication Conference and Exhibit, OFC,* pp. 222-224, 2002.

[92] I. de Miguel, E. Kozlovski et P. Bayvel, «Provision of end-to-end delay guarantees in wavelength-routed optical burst-switched networks», *Next Generation Optical Network Design and Modelling,* pp. 85-100, 2003.

[93] I. Saniee et I. Widjaja, «A new optical network architecture that exploits joint time and wavelength interleaving», *Optical Fiber Communication Conference (OFC),* 2004.

[94] K. Ross, N. Bambos, K. Kumaran, I. Saniee et I. Widjaja, «Scheduling bursts in time-domain wavelength interleaved networks», *Selected Areas in Communications, IEEE Journal on,* vol. 21, n° 9, pp. 1441-1451, 2003.

[95] I. Widjaja et I. Saniee, «Simplified layering and flexible bandwidth with TWIN», *Proceedings of the ACM SIGCOMM workshop on future directions in network architecture,* pp. 13-20, 2004.

[96] A. Brzezinski, I. Saniee, I. Widjaja et E. Modiano, «Flow control and congestion management for distributed scheduling of burst transmissions in time-domain wavelength interleaved networks», *Optical Fiber Communication Conference (OFC),* p. 3, 2005.

[97] G. Cazzaniga, C. Hermsmeyer, I. Saniee et I. Widjaja, «A New Perspective on Burst-Switched Optical Networks», *Bell Labs Technical Journal,* vol. 18, n° 3, pp. 111-131, 2013.

[98] C. Nuzman et I. Widjaja, «Time-domain Wavelength Interleaved Networking with Wavelength Reuse», *Annual IEEE International Conference on Computer Communications (INFOCOM),* vol. 6, 2006.

[99] C. Nuzman et I. Widjaja, «Design and performance evaluation of scalable TWIN networks», *Workshop on High Performance Switching and Routing (HPSR),* pp. 152-156, 2005.

[100] Y. Su, I. Widjaja, H. He, X. Xu, Y. Tian, J. Gao et T. Ye, «Demonstration of a Time-domain Wavelength Interleaved Network prototype without optical buffers and fast switches in the core nodes», *Optical Fiber Communication Conference (OFC),* pp. 1-3, 2007.

[101] A. Ahmad, A. Bianco, E. Bonetto, D. Cuda, G. Castillo et F. Neri, «Power-aware logical topology design heuristics in wavelength-routing networks», *15th International Conference on Optical Network Design and Modeling (ONDM),* pp. 1-6, 2011.

[102] O. Community, «OMNet++», [On line]. Available: http://www.omnetpp.org/.

[103] M. R. Garey et D. S. Johnson, Computers and Intractability: A Guide to the Theory of NP-completeness, WH Freeman and Company, New York, 1979.

[104] M. R. Garey et D. S. Johnson, «Complexity results for multiprocessor scheduling under resource constraints», *SIAM Journal on Computing,* vol. 4, n° 4, pp. 397-411, 1975.

[105] J. Cao, W. Cleveland, D. Lin et D. Sun, «Internet traffic tends toward Poisson and independent as the load increases», *LECTURE NOTES IN STATISTICS-NEW YORK-SPRINGER VERLAG-,* pp. 83-110, 2003.

[106] W. Simpson, «PPP in HDLC-like Framing», Network Working Group, RFC1662, 1994.

[107] BTI Systems, «Seven requirement for an all-packet mobile backhaul network bti systems», 2012.

[108] Y. Su, I. Widjaja, H. He, X. Xu, Y. Tian, J. Gao et T. Ye, «Demonstration of a Time-domain Wavelength Interleaved Network prototype without optical buffers and fast switches in the core nodes», *Conference on Optical Fiber Communication/ National Fiber Optic Engineers Conference (OFC/NFOEC 2007) ,* pp. 1-3, 2007.

[109] Celtic Group, «SASER-Savenet», [On line]. Available: http://www.celtic-initiative.org/Projects/Celtic-Plus-Projects/2011/SASER/SASER-a-SaveNet/saser-a-default.asp. [Access: January 2014].

[110] Hewlett-Packard, «HP Desktops», [On line]. Available: http://www8.hp.com. [Access: December 2013].

[111] National Instruments, «NI PXIe-1082 User Manual», February 2010. [On line]. Available: http://www.ni.com/pdf/manuals/372752b.pdf. [Access: January 2014].

[112] National Instruments, «NI PXI-8108 User Manual», March 2009. [On line]. Available: http://www.ni.com/pdf/manuals/372561d.pdf. [Access: January 2014].

[113] National Instruments, «NI FlexRIO FPGA Module Specifications», November 2009. [On line]. Available: http://www.ni.com/pdf/manuals/372525d.pdf. [Access: January 2014].

[114] National Instruments, «NI 6581 Specifications», May 2009. [On line]. Available: http://www.ni.com/pdf/manuals/372629b.pdf. [Access: January 2014].

[115] National Instruments, «NI SMB-2163», April 2004. [On line]. Available: http://www.ni.com/pdf/manuals/323660c.pdf. [Access: January 2014].

[116] Anritsu, «MT1810A 4 Slot Chassis Operation Manual», August 2013. [On line]. Available: http://www.anritsu.com/en-AU/Downloads/Manuals/Operations-

Manual/DWL9545.aspx. [Access: January 2014].

[117] Anritsu, «MU181020A 12.5 Gbit/s PPG MU181020B 14 Gbit/s PPG Operation Manual», November 2013. [On line]. Available: http://www.anritsu.com/en-AU/Downloads/Manuals/Operations-Manual/DWL9288.aspx. [Access: January 2014].

[118] Finisar Corporation, «CW Tunable Laser – Butterfly Package S7500», December 2011. [On line]. Available: http://www.finisar.com/sites/default/files/pdf/7500001-S7500%20CW%20Tunable%20Laser%20Spec-RevB.pdf. [Access: January 2014].

[119] R. Laroy, G. Morthier, R. Baets, G. Scarlet et J. Wesstrom, «Characteristics of the new Modulated Grating Y laser (MG-Y) for future WDM networks», *IEEE/LEOS Benelux Annual Symposium,* pp. 55-57, 2003.

[120] Finisar Corporation, «Controlling the S7500 CW Tunable Laser (AN-2095)», Sunnyvale, California, 2011.

[121] J. M. Fabrega, B. Schrenk, F. Bo, J. A. Lazaro, M. Forzati, P. Rigole et J. Prat, «Modulated Grating Y-Structure Tunable Laser for-Routed Networks and Optical Access», *Selected Topics in Quantum Electronics, IEEE Journal of,* vol. 17, n° 6, pp. 1542-1551, 2011.

[122] Y. Matsui, D. Mahgerefteh, X. Zheng, X. Ye, K. McCallion, H. Xu, M. Deutsch, R. Lewen, J.-O. Wesstrom, R. Schatz et P.-J. Rigole, «Widely Tuneable Modulated Grating V-Branch Chirp Managed Laser», *35th European Conference on Optical Communication (ECOC 2009),* pp. 1-2, 2009.

[123] Photline technologies, «Modulators : MX-LN-10», February 2012. [On line]. Available: http://www.photline.com/product/view/33/. [Access: January 2014].

[124] E. L. Wooten, K. M. Kissa, A. Yi-Yan, E. J. Murphy, D. A. Lafaw, P. F. Hallemeier, D. Maack, D. V. Attanasio, D. J. Fritz, G. J. D. McBrien et E. Bossi, «A review of lithium niobate modulators for fiber-optic communications systems», *Selected Topics in Quantum Electronics, IEEE Journal of,* n° 1, 2000.

[125] Photline Technologies, «RF drivers and modules : DR-DG-10-MO-NRZ», [On line]. Available: http://www.photline.com/product/view/1/. [Access: January 2014].

[126] Agilent Technologies, April 2000. [On line]. Available: http://cp.literature.agilent.com/litweb/pdf/81110-91020.pdf. [Access: January 2014].

[127] T. LeCroy, «WaveSurfer MXs-B and MSO MXs-B oscilloscopes», July 2012. [On line]. Available: http://teledynelecroy.com/oscilloscope/oscilloscopeseries.aspx?mseries=339. [Access: January 2014].

[128] Anritsu, «MS9740A Optical Spectrum Analyzer Operation Manual», July 2013. [On line]. Available: http://www.anritsu.com/en-US/Downloads/Manuals/Operations-Manual/DWL9999.aspx. [Access: January 2014].

[129] A. Castro, L. Velasco, M. Ruiz, M. Klinkowski, J. P. Fernandez-Palacios et D. Careglio, «Dynamic routing and spectrum (re) allocation in future flexgrid optical networks», *Computer Networks,* vol. 56, n° 12, pp. 2869-2883, 2012.

[130] C. Alaettinoglu, «Software Defined Network», Packet Design, Santa Clara, USA, 2013.

Résumé

Les réseaux d'opérateur du futur devront supporter des interfaces à haut débit et des besoins dynamiques de bande passante. Dans ce contexte, la commutation en sous-longueur d'onde pourrait remplacer avantageusement la commutation de circuit optique (OCS) afin d'apporter de la flexibilité à la couche optique. Dans ce contexte, TWIN (Time-domain Wavelength Interleaved Network) est une solution prometteuse. Elle fournit une commutation en rafale optique, sans perte, avec des nœuds intermédiaires passifs fonctionnant uniquement dans la couche optique, sur une topologie maillée.

L'utilisation de tels nœuds intermédiaires permet d'espérer une consommation électrique fortement réduite. Par contre, cette simplicité impose de recourir à un plan de commande complexe qui doit éviter les collisions entre rafales optiques.

Dans cette thèse, nous proposons plusieurs mises en œuvre du plan de commande et du fonctionnement du plan de données associé. L'évaluation des mécanismes proposés est conduite par simulation. Un banc expérimental a également été développé afin de montrer la faisabilité de ces technologies.

Mots-clés : Réseaux optiques, Commutation de rafales optiques, TWIN

Abstract

Future networks will have to support very high bitrate interfaces and to ensure dynamic bandwidth provisioning in order to deal with increasing and time-varying traffic demands. In this context, a sub-wavelength switching paradigm may be more appropriate than the currently deployed wavelength switching solutions as it brings flexibility in the optical layer. Time-domain Wavelength Interleaved Networking (TWIN) is a promising solution that provides lossless sub-wavelength switching using optical bursts.

The network topology is meshed with passive intermediate nodes operating only in the optical layer. A dedicated wavelength is assigned to each destination and each source can emit bursts to each destination thanks to a tunable laser. The use of passive intermediate node could reduce the energy consumption compared with electronic switches. However, the optical transparency requires a robust control plane in order to avoid burst contention.

Through this thesis, we propose several mechanisms to implement the control plane and its associated data plane. Simulation studies are carried out to assess the performance of our algorithms. Furthermore, an experimental test-bed is designed to prove the feasibility of these technologies.

Keywords : Optical networks, Optical burst switching, TWIN

TELECOM Bretagne

n° d'ordre : 2014telb0317

Télécom Bretagne

Technopôle Brest-Iroise - CS 83818 - 29238 Brest Cedex 3

Tél : + 33(0) 29 00 11 11 - Fax : + 33(0) 29 00 10 00